CounterStress: Enhancing Stress Coping Planning through Counterfactual Explanations in Personal Informatics

Gyuwon Jung
School of Computing
KAIST
Daejeon, Republic of Korea
gwjung@kaist.ac.kr

Uichin Lee*
School of Computing
KAIST
Daejeon, Republic of Korea
uclee@kaist.edu

Abstract

Personal informatics (PI) systems have been utilized to help individuals manage health issues such as stress by leveraging insights from self-tracking data. However, PI users may struggle to develop effective coping strategies because factors influencing stress are often difficult to change in practice, and multiple factors can contribute to stress simultaneously. In this study, we introduce CounterStress, a PI system designed to assist users in identifying contextual changes needed to address high-stress situations. CounterStress employs counterfactual explanations to identify and suggest alternative contextual changes, offering users actionable strategies to achieve a desired state. We conducted both lab-based and field user studies with 12 participants to evaluate the system's usability and applicability, focusing on the benefits of counterfactual-based coping strategies, how users select viable strategies, and their real-world applications. Based on our findings, we discuss design implications for effectively leveraging counterfactuals in PI systems to support users' stress-coping planning.

CCS Concepts

• **Human-centered computing → Ubiquitous and mobile computing design and evaluation methods**; **User studies**.

Keywords

Personal Informatics, Counterfactual Explanation, Coping Planning, Stress, Mental Health

ACM Reference Format:
Gyuwon Jung and Uichin Lee. 2025. CounterStress: Enhancing Stress Coping Planning through Counterfactual Explanations in Personal Informatics. In *CHI Conference on Human Factors in Computing Systems (CHI '25), April 26–May 01, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 20 pages. https://doi.org/10.1145/3706598.3713730

1 Introduction

The collection and reflection of personal data have become integral to daily life [4, 15]. Individuals typically track metrics, such as fluctuations in stress levels throughout the day, using data collected from multiple mobile devices [26]. This information is used to identify areas for improvement in health and well-being [70, 72]. Within the domain of Human-Computer Interaction (HCI), these practices are encapsulated under the concept of *Personal Informatics (PI)* [59], and prior studies have proposed models to explain the behavioral changes in users via PI systems [30, 60, 77].

PI users primarily engage in self-tracking for health management, seeking insights by analyzing relationships between various metrics [18]. Prior PI systems have shown how physical activity, daily schedules, and environments influence health indicators such as mood, weight, and sleep quality [5, 61]. They have also guided users through experiments to identify factors affecting health issues, such as sleep [23] and digestion [48]. Moreover, PI systems have evolved to support stress management by collecting stress-related data [67, 83], assessing stress levels [42, 87], and delivering results to users [52, 81]. Recent studies have employed causal inference to identify contextual stressors [45], predictive models to anticipate stressful events [58], and interventions to mitigate stress [51].

Despite these advancements, existing PI systems have limited support for coping strategies. Although they increase self-awareness of stressful situations, they provide little guidance on actionable coping strategies [21, 82]. Simply informing users about stressors may be insufficient, especially given real-world constraints where users cannot avoid certain stressors [49, 58]. For instance, students may recognize studying as a major stressor, yet avoiding it is unrealistic and not a practical coping strategy. According to Lazarus and Folkman [55], coping involves both identifying and managing stressors, highlighting the need for feasible and personalized coping strategies [2]. Even if users have some control over stressors, managing stress remains challenging due to the complex interplay of multiple factors [45]. Furthermore, existing interventions, such as meditation or stretching, often fail to account for users' specific contexts [43]. While these interventions may be beneficial, they lack the personal relevance needed to address individuals' stress effectively. However, by fully leveraging data collected by users, PI systems can bridge this gap by offering more targeted and context-aware coping strategies [51, 88]. Such strategies would empower users by facilitating self-reflection and equipping them with practical methods to manage their stress effectively [73].

To address these limitations, we propose *CounterStress,* a PI system as a mobile application, designed to support personalized stress-coping planning. In this study, we adopt Lazarus and Folkman's definition of "stress" from their transactional model [55], which conceptualizes stress as "*the psychological and physiological response to perceived challenges or threats*" that exceed an individual's coping capacity. Our study focuses primarily on acute and contextual stress, triggered by factors such as location, activity, and social

*Corresponding author

This work is licensed under a Creative Commons Attribution 4.0 International License.
CHI '25, Yokohama, Japan
© 2025 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-1394-1/25/04
https://doi.org/10.1145/3706598.3713730

settings in daily life [56, 79, 87]. These types of stress are dynamic and vary among individuals, making them well-suited for analysis using personal data in PI systems.

CounterStress leverages counterfactual explanations [40, 84], a widely used method in eXplainable AI (XAI) that answers, "*what changes are necessary to achieve a desired outcome*." In this study, counterfactual explanations offer valuable insights into "*what actions users should take to reduce stress to specific levels in a given situation composed of various contextual factors*." For example, in the situation comprising [studying, library, alone, afternoon], the system may suggest changing the location (e.g., library → dormitory) or altering the social setting (e.g., alone → friends). These recommendations are derived by identifying contextual changes that reduce stress probability, as predicted by a machine learning (ML) model trained on users' historical data. This approach ensures that suggestions are tailored to the patterns and relationships identified in the user's past stress responses, enhancing their relevance and practical applicability. Additionally, by generating multiple counterfactuals, users can select coping strategies that best fit their situation, enabling CounterStress to provide a highly personalized approach to stress-coping planning.

This study primarily investigates the user experience of coping planning based on counterfactuals provided by CounterStress and addresses the following research questions. **RQ 1:** How do users perceive the suggested coping strategies based on counterfactuals? **RQ 2:** How do users explore and select coping strategies based on counterfactuals? and **RQ 3:** How do users apply coping strategies based on counterfactuals in real-world settings? To address these questions, we conducted two complementary user studies: a lab-based study to evaluate the usability of the system and a field study to assess its applicability in real-world settings. Both studies involved the same 12 participants, which allowed for a comprehensive evaluation of CounterStress. The results indicate that CounterStress enables users to engage in stress-coping planning without the need for complex data analysis, providing a range of strategies that users can explore to effectively reduce stress. We also examined how users applied the recommendations from CounterStress to manage stress in their daily lives, and the criteria they used to determine and select the most suitable coping strategies.

This study contributes to the field of PI and stress management in several ways. First, we apply the counterfactual explanation approach in PI systems to assist users in planning effective coping strategies for stress management. Through this, we identify how PI users explore, assess, and utilize counterfactual-based coping strategies to address stressful situations. We also discuss the challenges associated with designing such systems and propose design considerations for offering feasible and effective coping strategies through the use of counterfactuals in PI systems.

2 Related Work

2.1 Stress and Approaches to Coping

Stress is known as a state of imbalance induced by internal/external forces that disrupt an individual's ability to maintain stability [19]. This imbalance triggers both physical and mental adaptive responses aimed at restoring equilibrium when the disturbance exceeds a specific threshold. Hans Selye, a pioneer in stress research, defined stress as "*the nonspecific response of the body to any demand, whether it is caused by, or results in, pleasant or unpleasant conditions*" [80].

Lazarus and Folkman's transactional model of stress and coping provides a key theoretical framework for assessing the perception and management of stress by individuals [55]. Within this model, "*coping*" refers to the various cognitive and behavioral strategies employed by individuals to manage stressful situations [33]. Their model delineates two types of cognitive appraisal: primary appraisal, which assesses whether a situation is positive, irrelevant, or stressful concerning an individual's well-being, and secondary appraisal, which assesses available resources or abilities to cope with the situation. Lazarus and Folkman distinguish two primary types of coping strategies: problem-focused coping, which involves directly addressing the cause of stress, and emotion-focused coping, which aims at regulating the emotions triggered by the stressful situation [34, 57]. Additionally, Schwarzer proposed four types of coping based on temporal factors (past vs. future) and the certainty of events (uncertain vs. certain) [75]. Among these, anticipatory and preventive coping involves preparing for potential future threats. Such strategies may include proactive problem-solving or securing resources in advance.

Drawing inspiration from these foundational studies, our research aimed to design a system that supports individuals in planning coping strategies for effective stress management. We assumed that various contextual factors experienced by individuals every day may serve as potential stressors. The proposed system in this study is designed to identify these stressors and deliver coping strategies based on collected user data.

2.2 Personal Informatics for Stress Management

The collection of personal data through various devices has become increasingly prevalent in daily life, and individuals reflect on and utilize this data for a wide range of purposes [28]. Li et al. [59] previously referred to systems supporting such practices as 'personal informatics (PI),' and proposed a model that identifies the barriers faced by PI users and provides suggestions for overcoming these challenges. PI systems are typically designed to help users change their behavior, and prior studies have analyzed the usage behavior of those systems by investigating the motivations and methods through which users engage in self-tracking [30, 60, 77].

One of the primary objectives of using PI systems is to enhance health and well-being [35, 65]. Choe et al. [18] revealed that PI users utilize self-tracking data to monitor their condition, identify influencing factors, and plan coping strategies, with the ultimate goal of improving their health state. Moreover, users seek to explore the collected data as well as to gain insights into the temporal progression of specific metrics, their distribution patterns, and their interrelationships among different metrics [16, 17, 49]. Consequently, various PI systems have been proposed for multiple health scenarios, such as physical activity tracking [53], nutrition monitoring [63], and menstrual cycle tracking [29]. These systems are designed to support users in reflecting on their self-tracking data and derive meaningful insights [15]. With these systems, users can develop a more nuanced understanding of their health across different timeframes, including the past, present, and future, thereby enabling them to take informed actions for achieving their desired health outcomes [4, 6, 74].

PI systems have been studied extensively in the field of mental health, particularly in stress management. Approaches leveraging mobile device sensor data [1, 67] and user self-reports [24, 83] have been primarily employed to assess stress levels. Wang et al. [87] collected data on activities, conversations, sleep, and location through smartphone sensors while concurrently gathering stress-related data via self-reports. Likewise, Hovsepian et al. [42] developed a predictive model for stress by leveraging data from multiple wearable sensors. Prior studies have also explored methods for stress detection based on physiological signals and behavioral data [13, 37].

In addition, extensive research has been conducted to assist users in monitoring and reflecting on their collected data to identify effective strategies for stress management [2]. For example, previous studies have proposed methods for visualizing various contextual data and stress levels to derive insights that inform stress interventions [52, 81]. In recent HCI research, Jung et al. [45] designed a system that presents contextual factors causally related to stress, whereas Lee et al. [58] and Kim et al. [51] proposed systems to predict future stress and provide interventions for stress management.

While existing studies have made significant progress in understanding stress factors, there remains a need for approaches that provide data-driven coping strategies tailored to the specific situations users encounter. When multiple factors influencing stress coexist, users often find it challenging to determine the necessary actions to reduce their stress levels [45]. Furthermore, practical constraints frequently prevent users from directly controlling stress-inducing factors [49, 58]. As a result, they are often left with generic solutions that do not align with their unique circumstances [43]. To address these limitations, this study proposes a PI system designed to generate and deliver stress-coping strategies specific to users' situations. The system aims to facilitate stress management by offering diverse alternatives that account for users' constraints while supporting flexible coping planning.

2.3 Counterfactual Explanations and Exploring Alternatives

As artificial intelligence (AI) continues to be integrated across diverse domains, the necessity for transparent explanations of trained models has become more critical. A significant challenge posed by many AI systems is that machine learning (ML)-based methods typically operate as "black boxes," hindering the ability to interpret and trust the predictions generated by these models [38].

XAI offers several approaches to address these issues by providing explanations for the results generated by ML models [27]. For example, SHapley Additive exPlanations [62] are employed to assess the contribution of each feature to a model's prediction, whereas Local Interpretable Model-agnostic Explanations [76] provide explanations based on simplified local models for specific predictions. In addition, methods such as the Partial Dependence Plot [36] and Individual Conditional Expectation [39] are frequently used to visualize the influence of specific features on predictions.

Counterfactual explanations represent another widely recognized approach in XAI that provides specific explanations about the modifications required to alter a predicted outcome generated by a model [40, 84]. Unlike other XAI methods, which primarily focus on analyzing why the model makes a particular decision, counterfactual explanations aim to identify the precise conditions required to modify that outcome. For instance, counterfactual explanations can be employed to determine the changes, such as job status, housing, and credit amount necessary for customers to improve their credit risk rating [22]. This method identifies the necessary changes in the feature values of a given instance to achieve the desired outcome, thereby providing multiple alternative scenarios (i.e., counterfactuals). By addressing the "what if" questions that humans naturally consider through counterfactual thinking [12], this approach facilitates more intuitive and human-friendly explanations.

A seminal contribution to the field of counterfactual explanation was made by Wachter et al. [85], who introduced a foundational method for generating meaningful counterfactuals. According to their approach, a counterfactual X' must satisfy two primary conditions: (1) the predicted outcome $f(X')$ by the classifier f should be as close as possible to the desired outcome Y', and (2) the counterfactual should be as similar as possible to the original instance X. These conditions guide the construction of a loss function, which is subsequently optimized to identify the appropriate counterfactual. Specifically, the second condition ensures that the generated counterfactuals minimize changes from the original instance, thereby maximizing the similarity between the counterfactual and the original instance.

Aligning with Wachter's approach, several other optimization-based algorithms for generating counterfactuals have been developed. These include, Diverse Counterfactual Explanations [68], Feasible and Actionable Counterfactual Explanations [71], Contrastive Explanation Method [25], and Multi-Objective Counterfactuals Explanation [22]. Although these algorithms differ in their specific methods, they all rely on optimization to balance the need for minimal changes in the input with the necessity of altering the model's prediction. Counterfactuals generated by these algorithms exhibit several key characteristics [40]. These include (1) **validity**, which ensures that the classification outcome differs from the original instance; (2) **similarity**, which guarantees that the counterfactual maintains the minimum possible distance from the original instance according to a given distance function; (3) **minimality**, which assesses whether the number of altered features is minimized; and (4) **plausibility**, which ensures that the counterfactual consists of realistic and feasible feature values.

Despite the significant potential of counterfactual explanations to simulate various what-if scenarios and provide actionable insights, their application in PI systems remains largely unexplored. As described above, existing PI systems have primarily focused on self-tracking and providing general insights, often lacking the ability to deliver tailored strategies for managing stress in complex, multi-contextual situations. To bridge this gap, our study integrates counterfactual explanations into a PI system specifically designed for stress-coping planning. By harnessing the strengths of counterfactual explanations, our approach identifies specific contextual changes that enable users to achieve their desired stress levels, even in scenarios involving multiple contextual factors. This facilitates the delivery of practical and actionable solutions that are applicable in real-world settings.

3 System Design

3.1 Design Rationale

Inspired by the concept of counterfactual explanations, we developed a PI system that supports users in planning strategies for stress coping. In brief, counterfactual explanations address the following questions: "Given a (factual) situation X and its outcome Y, what changes to X would be required to prevent Y from occurring (or to cause a different outcome Y')?" In other words, for a target situation X composed of n elements $\{x_1, x_2, \cdots, x_n\}$, and given that applying a classifier f results in $f(X) = Y$, the goal of counterfactual explanations is to determine the new situation $X' = \{x'_1, x'_2, \cdots, x'_n\}$ through minimal changes to the elements x_i of X so that $f(X') \neq Y$.

Our decision to incorporate counterfactual explanations stems from the underlying needs of PI users as identified in the existing literature. One of the key motivations for users to collect and reflect on their data in daily life is health management [18]. Users are particularly interested in identifying which factors are related to their health and how those factors influence their target health indicators. Therefore, they engage in *diagnostic tracking*, a form of self-tracking that involves collecting data to analyze the relationships between various factors [77]. Additionally, users frequently explore these relationships during the reflection phase through *dialogic reflection* [32], enabling them to uncover new factors that may impact their health and leverage these insights to take actions aimed at improving their well-being [17, 60].

Existing studies on PI systems also emphasize the necessity for users to manage their health and well-being. They have demonstrated correlations between various contextual factors and well-being metrics [5, 61], identified contextual factors that contribute to stress [45], and supported users in conducting self-experiments to assess the factors that influence their health [23, 48]. These studies indicate that users need to identify factors affecting their health based on self-tracking data and adjust these factors strategically to achieve their desired health objectives.

Therefore, we determined that counterfactual explanations could provide a valuable approach to supporting PI users. When utilizing counterfactuals to derive potential modifications for stress management, we anticipate several key advantages. First, counterfactuals enable users to predict the context changes required to effectively alleviate their stress. Particularly, counterfactuals facilitate users to target specific situations of interest, thereby providing information on the contextual factors that need to be altered and how these changes should be implemented. Based on these counterfactuals, users can develop tailored stress-coping strategies for each unique situation comprising different combinations of contextual factors.

Additionally, users can review multiple counterfactuals and select appropriate counterfactuals. Counterfactual explanations typically yield several counterfactuals in a specific situation [40]. Consequently, multiple coping strategies can be presented and users may select suitable strategies by assessing aspects such as the feasibility of the proposed counterfactuals. Considering these characteristics of counterfactual explanations, we derived counterfactual-based coping strategies aimed at assisting users in performing stress-coping planning with ease.

3.2 Counterfactual Explanations

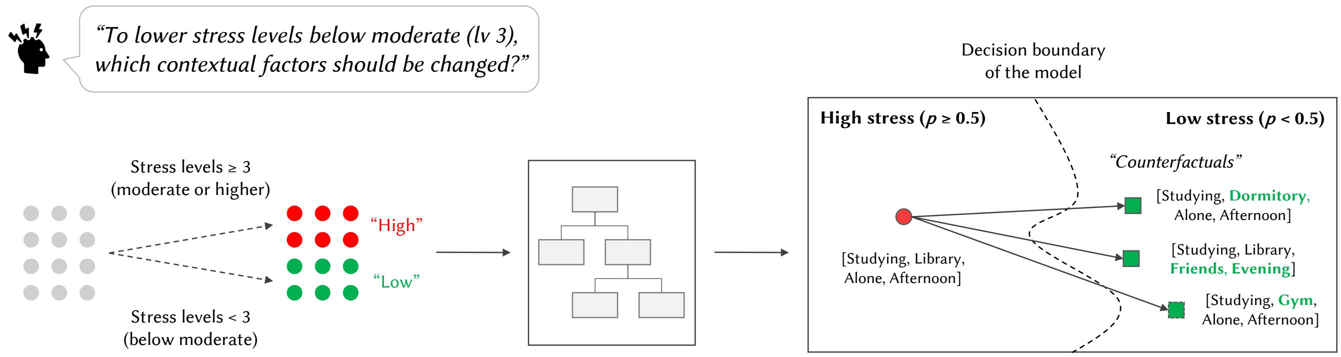
3.2.1 Generating Counterfactuals from Self-Tracking Data. We explain how counterfactuals were derived for specific situations from self-tracking data to recommend stress-coping strategies, as outlined in Figure 1. Hereafter, a **situation** refers to a combination of the four context types: [activity, location, social setting, time] (e.g., [studying, library, alone, afternoon]). In this study, each data sample we use (i.e., an individual record) consists of a situation and its associated stress level. Our objective was to identify the contextual changes necessary to reduce stress in each situation.

To simplify the analysis, we initially binarized the stress levels collected on a 5-point Likert scale (1: no, 2: mild, 3: moderate, 4: high, 5: severe) into 'high' and 'low.' In this study, our goal was not to eliminate stress entirely but rather to reduce it to an acceptable level. From this perspective, mild stress (level 2) was considered a manageable level of discomfort, typical of everyday life, and not necessarily requiring additional intervention. Participants were also guided before data collection to interpret mild stress as a state where stress is present but not particularly bothersome, reflecting a natural part of daily experiences.

As a result, we categorized levels 1 (no stress) and 2 (mild stress) as 'low stress' while levels 3 (moderate stress) and above were classified as 'high stress.' Among various approaches for binarizing Likert scale stress data [89], we adopted the common practice of using the midpoint value (i.e., moderate stress) as a binarization threshold, consistent with prior studies [8, 47]. Thus, we classified moderate stress as 'high stress' to emphasize the need for coping in such cases. Consequently, our approach focuses on answering the question: "In situations with moderate or higher stress, what contextual changes are necessary to reduce stress to lower levels?"

Next, we built a classification model using Random Forest to estimate the probability of stress being classified as 'high' (p) based on the context combinations. Given that each individual may experience different contexts and stress levels, we generated a separate model for each of the 12 participants, yielding an average accuracy of 0.79 (SD: 0.10). Counterfactuals were derived by evaluating which contextual changes would reduce the probability of experiencing 'high' stress ($p \geq 0.5$) below the threshold ($p < 0.5$), indicating situations with stress levels below moderate. This approach ensures that the suggested changes are data-driven and tailored to the patterns observed in each participant's historical stress responses. Based on this approach, we designed a user flow where the system generates appropriate counterfactuals tailored to the situation selected by the user for exploration. For example, if the stress level of the target situation was classified as 'high' ($p \geq 0.5$), the system provided counterfactuals suggesting contextual changes to help users transition to a 'low' stress state.

When generating counterfactuals, we established the following exclusion criteria to avoid unnecessary or irrelevant counterfactuals. First, counterfactuals were excluded if they simply added changes in other context types from existing counterfactuals but did not increase the likelihood of being classified as a 'low' stress. This decision was based on the criterion that the changes required to generate a counterfactual should be minimized. In addition, counterfactuals were excluded if none of their contexts overlapped with those of the original target situation. Although such changes might



(a) Label the samples' stress state based on a specific threshold (e.g., moderate) (b) Build an ML model (e.g., Random Forest) (c) Generate counterfactuals (the green squares) for the target situation (the red circle)

Figure 1: The process of investigating stress-coping strategies based on counterfactual explanations.

lower stress levels, we deemed them irrelevant owing to the lack of common context. Moreover, we applied what we refer to as "contextual familiarity thresholding" to ensure the relevance of the counterfactuals. In this approach, counterfactuals were excluded if any combination of the three contexts out of four context types was not recorded. This threshold was set to filter out unrealistic context combinations while still allowing for situations wherein one context type was omitted. This approach provides users with novel but plausible coping strategies.

Figure 1 illustrates an example derived from a participant's data. The situation [studying, library, alone, afternoon] was predicted as high stress ($p \geq 0.5$), prompting the system to generate counterfactuals that could reduce stress to a low level ($p < 0.5$) based on the trained ML model. One suggestion, studying in the gym, reflects the participant's past behavior of cycling at the gym while taking a short online course, making this recommendation relevant to their historical data. Although counterfactuals are generated based on one's historical data, their feasibility or desirability can vary depending on individual preferences and circumstances.

3.2.2 Evaluating the Quality of Counterfactuals. As previously explained, multiple counterfactuals can be generated for a given situation using existing algorithms. When evaluating the quality of these counterfactuals, previous studies have considered factors such as the probability of achieving the desired outcome, similarity to the feature values of the target instance, and the realism of the combinations of feature values. However, studies on the criteria that users prioritize when providing counterfactual-based coping strategies for PI systems are limited. To address this gap, we aimed to investigate how users weigh each criterion when selecting suitable coping strategies from a set of counterfactuals, based on the criteria typically employed in various counterfactual explanation algorithms. To facilitate this analysis, we presented users with three metrics that describe the criteria for multiple counterfactuals: (1) high-stress probability, (2) number of context changes, and (3) historical frequency. Below, we outline how each metric was applied in our analysis.

High-stress probability (p): We provided the probability p that a given counterfactual would result in the stress level being classified as 'high.' According to our approach, all counterfactuals presented to the users resulted in p below 0.5, classifying them as 'low' stress. However, the value of p may vary within the range of 0 to 0.5, resulting in differences in the probability across different counterfactuals. For instance, the likelihood of experiencing 'low' stress increases as p approaches 0, and we assessed how this variation in p influenced users' choices of coping strategies.

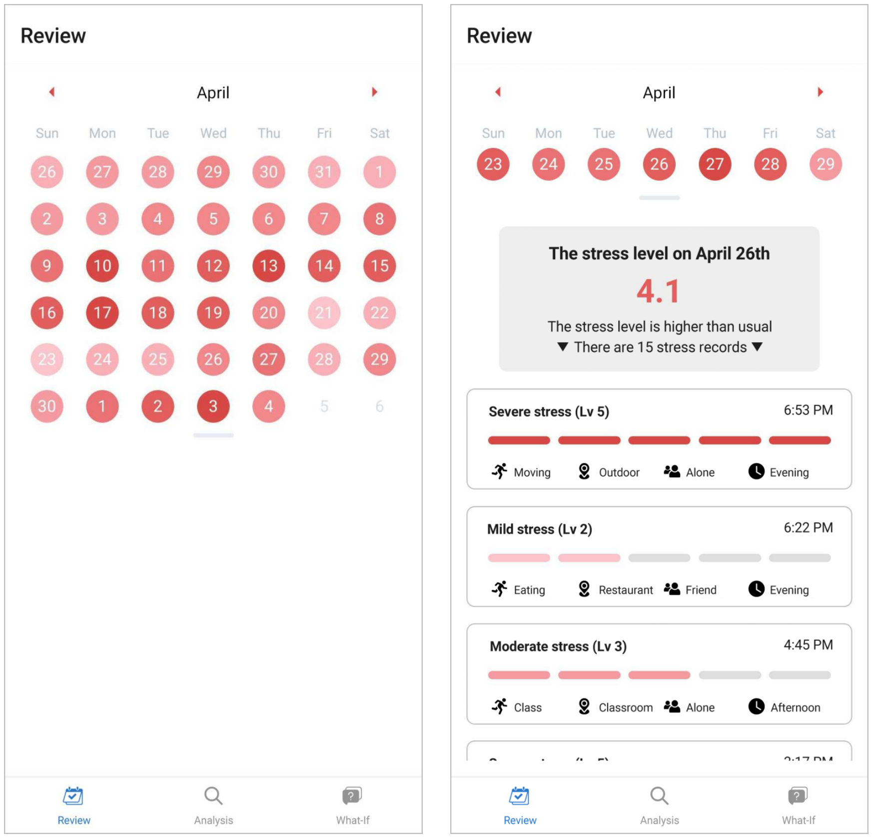
Number of context changes (n): Additionally, we presented the number of context changes n required to achieve a given counterfactual. Since changing all four context types was considered irrelevant and excluded from the counterfactuals, n could range from 1 to 3. An increase in n indicated that a greater number of contexts would need to be modified, which could result in a situation that diverges more significantly from the initial target situation. From this perspective, we analyzed how the magnitude of n influenced users' choices.

Historical frequency (r): Finally, we provided the frequency r representing how frequently the situation described by a counterfactual had occurred in the past. Even after applying exclusion criteria to eliminate irrelevant counterfactuals, the generated counterfactuals could include both situations that had been experienced before ($r > 0$) and those that had not ($r = 0$). In this sense, we explored how a zero or non-zero value of r influenced users' choices.

Although various other criteria for generating counterfactuals have been proposed in previous studies [40], our focus was primarily on these three factors that are typically employed to evaluate counterfactual-based decision making processes. In this analysis, we aimed to identify the crucial criteria for deriving counterfactuals that provide effective suggestions for users. Furthermore, we explored the design space for utilizing counterfactual explanations to PI systems based on the insights gained from the analysis.

3.3 CounterStress

We designed CounterStress, a PI system that provides coping strategies based on the counterfactuals generated through the process



(a) (b)

Figure 2: The *Review* screen. (a) It presents a monthly calendar with color-coded stress levels and (b) provides detailed records when users select a specific date.

outlined earlier. The primary objective of CounterStress was to assist users in (1) reviewing their self-tracking records, (2) identifying the relationships between contextual factors and stress, and (3) exploring counterfactuals that propose contextual changes to reduce stress in specific situations.

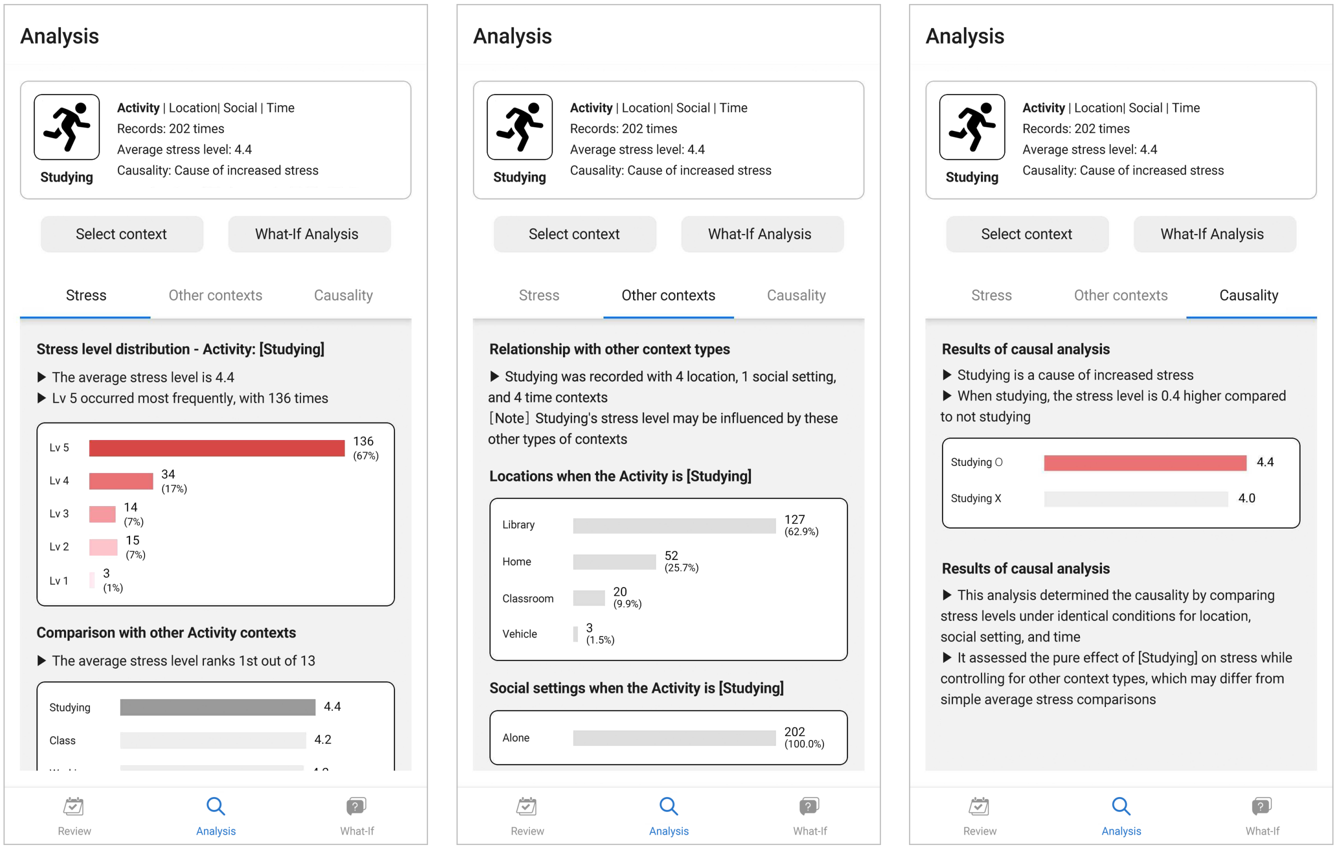
Our study builds on prior work [45] that designed a system to investigate stressors across diverse contexts. This closely related study focused on stress management by identifying which contextual factors should change, considering both correlation and causality. However, users had to develop coping strategies by themselves based on their understanding of each context. Therefore, we extended this study by delivering concrete strategies (i.e., required context changes) using counterfactuals, thereby offering more comprehensive and actionable insights for coping planning.

We followed an iterative design process with nine HCI researchers, conducting a mid-fidelity prototype test (N = 3) and a high-fidelity prototype test (N = 6). The decision to engage HCI researchers instead of non-researchers was motivated by their expertise in system usability and interface design. Their expert heuristic evaluations help identify critical usability issues and refine the system ahead of broader testing with end-users. Through this iterative process, several key improvements were made to CounterStress. For instance, the visualization of generated counterfactuals was enhanced to highlight differences in generation criteria between alternatives, making it easier for users to compare them. The researchers also suggested incorporating filtering features, allowing users to sort and prioritize counterfactual suggestions based on specific needs or preferences. These changes improved the system's usability and its ability to deliver actionable and personalized coping strategies.

Consequently, CounterStress consists of three main screens; 'Review,' 'Analysis,' and 'What-If,' and detailed descriptions of each of them are provided below.

3.3.1 Review. The Review screen (Figure 2) enables users to examine the stress records they collected. It aims to assist users in recalling past stress levels in various situations and identifying the contexts that may require coping planning.

The screen features a calendar view that visually represents monthly stress trends. In the calendar, each day is color-coded based on the average stress level, with darker reds signifying higher stress and lighter reds indicating lower stress. When selecting a specific day, users were presented with a summary of the day's average stress level and the total number of stress records. Below the summary, detailed records were presented in chronological



(a) (b) (c)

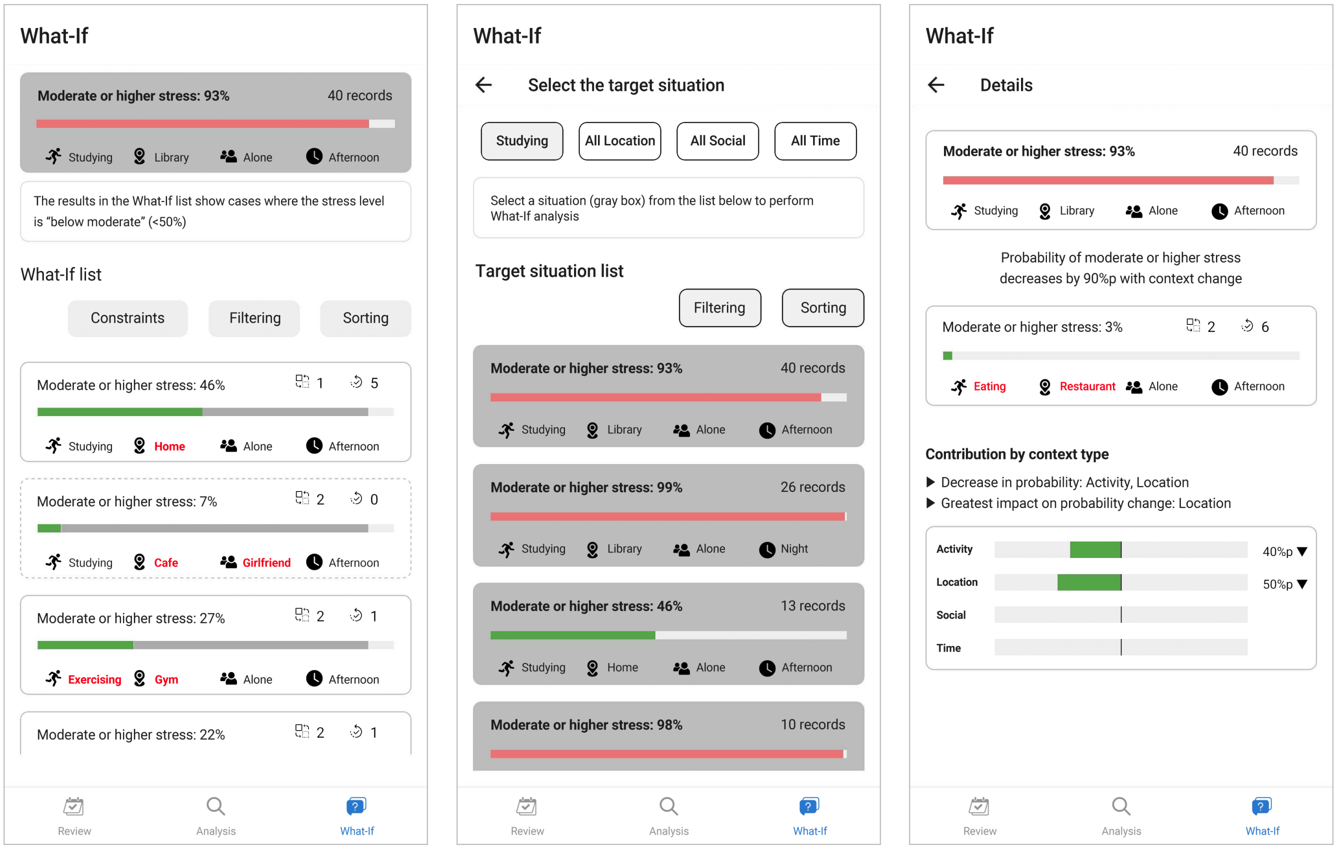
Figure 3: The *Analysis* screen. It provides information for each contextual factor and consists of three tabs: (a) *Stress*, which describes the distribution of stress levels, (b) *Other contexts*, which illustrates co-occurring contexts, and (c) *Causality*, which explains whether the selected context is causally related to stress levels.

order, including stress levels on a 5-point Likert scale and the four types of contexts.

3.3.2 Analysis. The Analysis screen (Figure 3) allows users to understand how individual contextual factors relate to stress. It provides insights into stress levels and identifies causal relationships in each context, aiding users in pinpointing contexts that may require further exploration for effective coping strategies. The screen begins with a summary of the selected context, detailing its type, frequency, average stress level, and causal relationships with the stress. This provides users with a quick overview before the detailed analysis. The screen is organized into three tabs: 'Stress,' 'Other Contexts,' and 'Causality.'

The Stress tab displays a bar chart of the stress level frequencies within the selected context, helping users understand the distribution of stress levels. Additionally, it compares the average stress level of the selected context with that of other contexts of the same type. The Other Contexts tab illustrates the distribution of context types that co-occurred with the selected context. For instance, if the selected context was the activity type 'studying,' this tab displayed information on the remaining contextual factors (i.e., location, time, and social setting) that accompanied the studying.

The Causality tab explains whether the selected context has a causal relationship with stress based on a quasi-experimental approach with coarsened exact matching [7, 44], as detailed in the prior work [46]. The process of causal inference can be summarized as follows. For instance, to investigate the causal relationship between studying and stress, the dataset is divided into a treated group (cases where the activity is 'studying') and a control group (cases where it is not). The remaining context types (i.e., location, social setting, and time) that may affect the causal relationship are treated as confounding variables. A matching process is then applied to pair treated and control samples with similar combinations of these confounding variables. Using these matched samples, treated and control groups are reconstructed, balancing the distribution of confounding variables between the two groups. This ensures that any difference in stress levels between the groups is solely attributable to the activity being 'studying.' If a causal relationship existed, the Causality tab displayed the extent to which stress level increased or decreased when the context was present. A brief explanation of



(a) (b) (c)

Figure 4: The *What-If* screen. It provides (a) counterfactual-based coping strategies for a selected situation, (b) a list of situations to be investigated, and (c) details on the contribution of each context type for the selected counterfactual.

the analytical methods employed and instructions for interpreting the results was also provided.

3.3.3 What-If. The What-If screen (Figure 4) enables users to review and explore counterfactual outcomes for the situations of interest. The target situation was displayed within a gray box at the top, showing the probability p of the stress level being classified as 'high,' along with a bar chart in red or green depending on whether p was above or below 50%. Also, the gray box included the contexts and the number of records associated with the target situation.

Below the target situation, the What-If list presents a series of counterfactuals in white boxes, each containing information similar to that of the target situation. The number of contextual changes (n) is displayed for each counterfactual, and the changes are highlighted in bold red. Moreover, counterfactuals with previous occurrences ($r > 0$) are indicated with solid borders, whereas those with no prior occurrence ($r = 0$) are outlined with dashed borders. To facilitate the identification of the desired counterfactuals, the users were provided with sorting, filtering, and constraint features. Sorting allows users to arrange counterfactuals based on the magnitude of the reduction in probability p or by the number of contextual changes n. Filtering enables users to specify ranges for p and n within the generated counterfactuals. The constraint feature empowers users to specify which contexts in the target situation should remain unchanged. For instance, if users constrained activity and location, CounterStress would generate counterfactuals that only modified social setting and time.

Users could modify the target situation by tapping the gray box, which displays only previously experienced situations ($r > 0$). By selecting a situation, users can explore the necessary contextual changes required to reduce stress. Following the selection of a target, users are returned to the What-If screen, where corresponding counterfactuals are displayed. Filtering and sorting tools were also available to refine the target selection. Additionally, users could directly specify which contexts to include in the target situation by selecting from each context type.

When users selected one of the counterfactuals from the What-If screen, they can view the contribution of each context type to the reduction of the probability p. This contribution is quantified using the Shapley value for each context type, and the results are visualized using a bar chart. This feature was provided to assist users in understanding the impact of changes in each context, particularly in cases where multiple contexts needed to be changed.

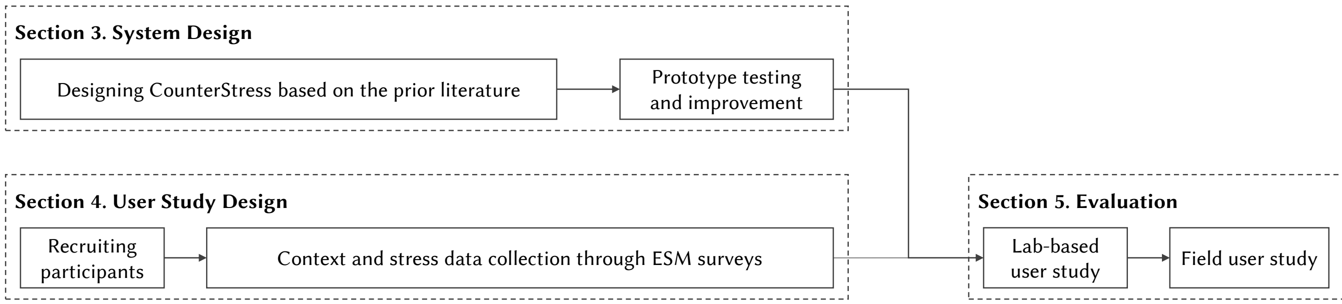


Figure 5: An overview of the study procedure and its corresponding sections.

4 User Study Design

4.1 Participants

We recruited participants through an online community and email at a large university to evaluate the user experience of CounterStress. As part of the recruitment, applicants were required to complete an online survey. Survey responses were reviewed to determine eligibility, and only those who met our predefined criteria were selected as final participants. First, the survey incorporated the Perceived Stress Scale (PSS) [20] to measure the level of stress participants experienced in their daily lives. Applicants with a PSS score of 12 or below (i.e., 'low stress' level) were excluded as they seldom experience moderate or higher levels of stress. We assumed that these individuals might have limited experience in exploring stress management strategies, which would make them less aligned with the aim of our study (i.e., stress-coping planning).

The survey also evaluated the applicants' level of interest in stress management (e.g., "I am interested in using my daily life data to plan stress-coping strategies.") and their ability to perform repetitive tasks over an extended period (e.g., "I am generally good at completing tasks that require consistent repetition.") on a 5-point Likert scale. To ensure sufficient self-reported data, we prioritized motivated applicants. Therefore, applicants who scored below 3 for either item were excluded. As a result, we recruited 12 participants (3 women, 9 men, age: M = 23.1, SD = 2.1) as illustrated in Table 1. These participants participated in both lab-based and field user studies, as described in the following sections.

Table 1: Participant demographics and selection attributes

Participant ID	Age	Gender	Stress Interest	Task Consistency
P01	24	M	5	5
P02	24	M	5	4
P03	25	M	5	4
P04	25	M	5	5
P05	24	F	4	4
P06	21	F	4	5
P07	21	M	5	4
P08	22	M	4	5
P09	23	M	5	4
P10	20	M	5	3
P11	27	M	5	4
P12	21	F	4	3

4.2 Study Procedure

During the study, we designed 'CounterStress' and explored its user experience as illustrated in Figure 5. The system was designed based on findings from existing literature in the fields of PI and counterfactual explanations, as outlined in Section 3. The participants collected the real-world data required to evaluate the proposed system by reporting their context and stress levels at specific intervals over six weeks. After the data collection phase ended, the self-reported data were used for the two user studies (details in Section 4.3).

The first study was a **lab-based user study** that primarily examined how participants used CounterStress in a lab setting and evaluated the system's data-driven insights. We started by explaining the background and purpose of the study to the participants and briefly introducing the features and information provided by CounterStress. The participants subsequently installed CounterStress on their smartphones and reviewed their data alongside the insights derived from it. Specifically, they were able to assess the changes necessary to reduce stress in particular situations based on the different contexts and stress levels they experienced.

Participants were allocated up to 40 minutes to explore CounterStress, during which they engaged with various pieces of information provided. After that, they completed the System Usability Scale (SUS) [11] to evaluate the usability from the perspective of supporting their stress-coping planning. For qualitative evaluation, we also conducted semi-structured interviews, focusing primarily on participants' overall assessment of CounterStress and how they explored the information provided. We recorded the interview sessions with the participants' consent and transcribed them.

The second study was a **field user study** aimed at investigating how participants used CounterStress and applied the provided information in their daily lives. Participants began the field study the day after completing the lab-based study. This arrangement allowed them to immediately apply the knowledge and coping strategies they had explored during the lab-based study in their real-world contexts. By minimizing the time gap between the two studies, we ensured that the participants retained familiarity with the features and insights of CounterStress, thus facilitating a smoother transition to using the system in their daily lives.

We instructed participants to write a diary for one week to capture their everyday usage experiences with CounterStress. The diary method was selected owing to its ability to capture reflective thoughts and subjective experiences in natural environments [9].

This approach has been widely used in HCI research to gather detailed insights into user behavior and interaction with systems in real-world settings [3, 14]. Leveraging this method allowed us to gain deeper insight both into how the participants engaged with the system as well as the underlying motivations and broader context of their behavior, thereby providing rich qualitative insights into the real-world impact of CounterStress insights that are challenging to uncover via usage logs. Participants documented various episodes including instances when they revisited CounterStress, applied the coping strategies suggested by the system, or gained new insights. The participants submitted diaries after the one-week study period. The details of the data analysis process are described in Section 4.4.

Participants were compensated a total of 90 USD for participating in data collection and the two user studies. This research was approved by the Institutional Review Board (IRB) of the researchers' affiliated institution, and all participants provided written consent to participate in the study.

4.3 Self-Tracking Data Collection

As previously discussed, we instructed the participants to collect their own data to evaluate their CounterStress. Given that CounterStress is a PI system, we determined that evaluating coping strategies derived from the participants' own data would be more suitable than using data from external sources, which may lack relevance, or synthesized data, which do not reflect real-world experiences. Over a six-week period, participants collected data comprising contexts and stress levels through the Experience Sampling Method (ESM) [54] via a separate mobile application we provided.

The use of ESM to collect self-reported stress data is a well-established method in prior relevant studies [42, 47, 66, 87]. Following these studies, we sent ESM surveys at intervals of approximately 1 to 1.5 hours to record the real-time stress levels of participants. Several measures were implemented to minimize the potential impact of this data collection method on the participants' stress. Participants were not sent additional follow-up reminders if they missed the ESM survey, thereby ensuring that responses were natural and not influenced by external pressure. The ESM surveys were designed as multiple-choice questionnaires, enabling participants to quickly input their context and stress levels with minimal cognitive effort. This approach ensured that participants could report their stress immediately without the need for extensive reflection. In addition, participants could customize the time windows for receiving ESM survey notifications, increasing the likelihood of responses while preventing disruptions such as late-night notifications.

Through ESM surveys, we collected data on three contextual factors (i.e., activity, location, and social setting) up until the time of the response and participants' stress levels when responding, rated on a 5-point Likert scale (1: no stress, 5: severe stress). To streamline the data collection, we provided context options for each type, referring to a relevant prior study [50], while allowing them to input their context manually if needed. In most cases, the participants selected options provided to report their context. When participants manually reported a context, researchers categorized it as the most similar predefined option for use in later data analysis. This decision was made to handle rare cases while still ensuring that as much data as possible were gathered. Note that the time context was not directly collected; instead, labels were generated based on the time at which the data were gathered. The context collected is summarized in Table 2, and participants reported an average of 558.4 ESM surveys (SD: 144.2) during data collection. Also, the stress levels reported by participants averaged 2.5 (SD: 0.7) on a 5-point Likert scale.

Table 2: The list of contextual factors collected through ESM surveys in this study.

Context Type	Context Items
Activity	Studying, Class, Resting, Working, Research, Meeting, Exercising, Eating, Social Activities, Drinking, Leisure Activities, Club Activities, Moving, Waiting, Preparing, Others
Location	Dormitory, Home, Classroom, Library, Laboratory, Workplace, Restaurant, Cafe, Pub, Store, Gym, Club Room, Vehicle, Outdoors, Leisure Facility, Others
Social Setting	Alone, Family, Friend, Girlfriend/Boyfriend, Roommate, Colleague, Professor, Others

4.4 User Study Data Analysis

To gain deeper insights into user experiences of CounterStress, we conducted a qualitative analysis of the two user studies (Section 4.2). Following established methods [10], two researchers independently reviewed the data to ensure diverse perspectives and enhance the reliability of the findings. Each researcher independently reviewed all interview transcripts multiple times to develop a deep understanding of the data and conducted open coding to identify the key concepts. This independent coding process ensured that a variety of insights were captured. Following the initial coding, the researchers held collaborative discussions to compare codes, resolve discrepancies, and refine the coding scheme. Using this refined scheme, the codes were organized into broader themes, representing patterns across the data. The researchers then engaged in iterative discussions to review and refine these broader themes, and reached a consensus on the final themes. Affinity diagramming was employed to visualize the relationships between these final themes, providing a comprehensive understanding of the participants' experiences with the system.

Qualitative analyses were conducted separately for the lab-based and field-based user studies. For the lab-based study, the aforementioned analysis process was applied to the interview data, focusing on the immediate reactions of the participants and the exploration of CounterStress in a controlled environment. For the field-based study, the same approach was applied to diary records that captured participants' reflections and nuanced stress management practices using CounterStress across diverse real-world situations. Although the same analytical approach was applied in both studies, the analyses were conducted independently to reflect the specific contexts in which the data were collected, allowing for a more comprehensive understanding of diverse user experiences.

Table 3: Key findings from the evaluation of CounterStress, organized by research questions and corresponding user studies

Section	Research Question	Source	Key Findings
5.1	RQ1: How do users perceive suggested coping strategies based on counterfactuals?	Lab-based user study	(1) Enables users to simulate the impact of contextual changes on stress levels without requiring complex data analysis (2) Provides detailed and practical coping strategies to help reduce stress in specific situations (3) Offers a variety of coping strategies and allows users to compare their impact on stress levels
5.2	RQ2: How do users explore and select coping strategies based on counterfactuals?	Lab-based user study	(1) Determined target situations for exploring coping strategies based on high-stress probabilities and frequency of occurrence (2) Applied constraints for unavoidable contexts and explored counterfactual scenarios by modifying other contextual factors (3) Evaluated coping strategies using counterfactual generation criteria, with historical frequency identified as the most critical
5.3	RQ3: How do users apply coping strategies based on counterfactuals in the real-world settings?	Field user study	(1) Utilized CounterStress's coping strategies to plan upcoming activities, manage stress in real-time, and reflect on the day (2) Made contextual changes based on CounterStress's suggestions, influencing their behaviors and thoughts on stress management

5 Evaluation

We investigated our research questions through a lab-based user study and a field user study, as described in the previous section. As summarized in Table 3, the lab-based user study provided insights into RQ1 (users' perceptions of counterfactual-based coping strategies) and RQ2 (how users explore and select such strategies). Meanwhile, the field user study addressed RQ3 (how users apply counterfactual-based coping strategies in real-world settings).

5.1 RQ1: How Do Users Perceive Suggested Coping Strategies Based on Counterfactuals?

The lab-based user study indicated that the coping strategies suggested by CounterStress were useful for managing stressful situations in daily life. Participants particularly evaluated that coping strategies based on counterfactuals from the What-If analysis would be beneficial, considering both their data collection experience during this study and their prior use of PI systems. Our SUS survey with 12 participants yielded a mean score of 75.0 (SD: 15.6), indicating that the usability of CounterStress was 'good' for assisting users in planning stress coping. In this section, we present findings derived from interviews conducted during the lab-based study.

5.1.1 Enabling Coping Planning Without Complex Data Analysis. Participants were able to efficiently explore how changes in context within specific situations could impact stress levels using CounterStress. In the interviews, participants reflected on past difficulties in investigating how changes in each context affected stress levels in specific situations. P01 noted, "*I looked at what I was doing when my stress was the highest and tried to figure out what activity to switch to, but there were so many activities. If I had tried to change other factors too, well... it probably would've taken way too much effort.*" In addition, manual exploration has become more challenging, because stress levels vary among records containing the target context. P03 remarked, "*Even when I'm doing the same activity, my stress changes depending on when I do it or who I'm doing it with, so the situations I need to consider are getting more complicated.*"

Our results indicated that CounterStress allowed them to easily examine stress levels across various situations without performing complex data analysis, as P10 noted: "*The biggest advantage is that I can understand how to make positive changes in stressful situations simply by specifying the conditions I want to investigate, without analyzing everything myself.*" P03 said, "*I think CounterStress would be helpful for people who often feel stressed. Some might even go to counseling for it, but this could help them easily figure out what changes they need in their daily life.*" The participants appreciated the ability to simulate how stress might change without having to change contexts in real life, enabling them to assess situations that they were curious about. "*I like that CounterStress can predict whether a strategy I've never tried before will work for me or not. Otherwise, I'd have to try everything out blindly.*" (P04).

5.1.2 Providing Effective Coping Strategies for Reducing Stress. Participants were able to obtain detailed and practical coping strategies to reduce stress using CounterStress. They responded that the PI systems they had used in the past did not adequately support such features. Typically, these systems focused on displaying changes in health metrics (e.g., stress levels and sleep quality) over time. Therefore, while they were aware of their stress levels, they found it challenging to clearly determine actionable steps to manage their stress using their data. P09 noted, "*I often collect my data, but it doesn't really help me understand the reasons behind it in connection with other factors, so I haven't thought much about what I should change.*" Although some participants attempted coping planning, they questioned whether it was an effective approach, as P06 responded: "*With a typical stress app, I'd check when my stress was at its peak and think, 'I should do the opposite.' But I wouldn't be sure if just changing my activity would really reduce my stress.*"

However, CounterStress provided participants with specific recommendations for contextual changes in specific situations, enabling them to develop actionable coping strategies. "*Evaluating things by changing conditions one by one like this would help me create concrete strategies to lower stress effectively*" (P06). Particularly, participants appreciated the ability to observe how changes in other contexts affected stress levels while keeping a specific context fixed, enabling them to plan more realistic coping strategies. "*When it comes to studying or doing assignments, I can't avoid them, but I found that changing the time or who I'm with can lower my stress. So, I realized I could manage my stress more flexibly with these insights*" (P05). Additionally, participants reported that *the fact that the counterfactual scenarios were generated based on the user's actual data increased the reliability of the analysis results* (P05, P06, P10). However, there was also feedback that additional explanations were needed for the reliability of situations that were untried (P08).

5.1.3 Allowing Exploration of Various Coping Strategies. Participants positively evaluated CounterStress for its ability to provide multiple coping strategies tailored to specific situations. P09 said, "*It shows multiple strategies, allowing me to choose the ones that seem the most suitable for the given situation. I like that it offers various options to apply based on the situation.*" P04 also highlighted the availability of multiple strategies, "*Because I have a set routine, I'd probably rely on my past experiences when looking for coping strategies in the data. There might be better, new strategies out there, but finding them in the data doesn't seem easy without systems like this.*"

Furthermore, the ability to compare different coping approaches provided helped participants understand how different contexts interact and influence changes in stress levels. P12 responded, "*Even in similar situations, like studying with friends in the club room, I noticed how much my stress changes depending on the time. I realized small context changes can have a bigger impact than I expected.*" This experience motivated them to explore new approaches that they had not previously considered.

5.2 RQ2: How Do Users Explore and Select Coping Strategies Based on Counterfactuals?

In this section, we explore how participants explored and selected counterfactual-based coping strategies using CounterStress in the lab-based user study. The findings highlight their processes of engaging with the data and analysis results, as well as their criteria for evaluating and selecting coping strategies.

5.2.1 Selecting Target Situations for Exploring Coping Strategies. Participants primarily selected situations to investigate based on two criteria: those with a high probability of a 'high' stress state and those that occurred frequently. They focused on high-stress situations because their primary goal in using CounterStress was to reduce stress. They were particularly curious about stress reduction in unavoidable high-stress situations, prompting them to explore these cases more thoroughly. P07 mentioned, "*From the data and my own experience, it seemed like studying stressed me out the most, so I focused more on situations where I was studying. I was also curious about how I could study with less stress.*" There were also counterfactuals for low-stress situations (i.e., stress-increasing context changes), but participants were less interested in this information.

In addition, participants considered frequently occurring situations to be important because of their relevance and impact on their daily lives, as P08 noted. "*As a student, the contexts I typically experience are relatively fixed. So, when certain contexts occur frequently, they become major parts of my daily life and hold greater relevance to me.*" Additionally, they explained the importance of the situation's frequency in relation to stress levels, as P03 mentioned: "*If something only shows up once or twice with a high probability, I assume it's just a coincidence and move on. But if it happens often and has a high probability, I trust the results more and take a closer look.*" Furthermore, participants explored situations they were curious about regarding their impact on stress, as well as those in which they could make more changes.

Depending on the features provided by CounterStress, the participants determined the target situations for investigation based on either individual contexts or their combinations. They reviewed the information provided on the Review or Analysis screens and selected situations that included specific individual contexts as analysis targets. On the Review screen, they selected high-stress days from the calendar and identified commonly occurring contexts on those days. Then they chose to analyze situations in which these contexts were present, assuming that these contexts were likely to contribute to high stress. From the Analysis screen, they focused on contexts with high average stress levels or those identified as stressors, prioritizing those that were frequently recorded. Furthermore, the participants determined their analysis targets at the situational level (i.e., combinations of contexts) based on the information displayed on the What-If screen. They reviewed the stress information and frequency of occurrence for various situations and selected those that they wanted to explore as coping strategies.

5.2.2 Setting Constraints to Reflect Real-World Situations. Participants evaluated the 'constraint' feature as a key element when reviewing the counterfactual scenarios on the What-If screen. All participants utilized the constraint to fix certain contexts that they intended not to change and explored changes in the remaining contexts. The most commonly fixed contexts were of the 'activity' type, stemming from their primary question: "*In what situations should I perform these activities to experience less stress?*"

These activities typically represent essential tasks in daily life (e.g., studying, attending classes, or working), and they are unavoidable even if they are associated with high-stress states. Consequently, they explored coping strategies by modifying other context types (i.e., location, social setting, and time) while keeping the activity unchanged. Participants also fixed other types of contexts based on the characteristics of the activities and compared potential coping strategies within these constraints. For example, some activities had to occur at a specific time or place (e.g., *attending an in-person class in the morning* (P08)) or with certain people (e.g., *participating in a club activity with club members* (P06)).

As more constraints were added, the available options for coping strategies became very limited or even entirely unavailable. In such cases, participants gradually relaxed less critical constraints to secure more coping strategies. During this process, they suggested ways to increase the availability of coping strategies, such as *allowing the recording of more diverse types of information beyond the four context types* (P11) or *displaying potential strategies even if the probability is above 50%, as long as it represents an improvement over the initial situation* (P10, P12).

5.2.3 Assessing Coping Strategies Using Counterfactual Generation Criteria. When CounterStress provided several coping strategies based on counterfactuals, participants compared and determined which plan would be most appropriate for the given situation. They evaluated coping strategies from the perspective of three criteria considered in the counterfactual generation process: (1) high-stress probability, (2) number of context changes, and (3) historical frequency. We observed that these criteria served as auxiliary information, facilitating participants to easily identify differences between counterfactual situations.

Historical frequency: Among the three criteria for generating counterfactuals, most participants considered historical frequency the most important. They prioritized situations that they had experienced in the past, expecting these combinations to be more likely to occur. P02 mentioned, "*Considering my routine, I thought that situations with high frequency are likely to happen more often in the future.*" P06 also added, "*If a situation hasn't shown up in over a month of data collection, I think it's unlikely I'll encounter it in the future. It seems more practical to find ways within the situations I've already experienced.*"

The situations that participants had not previously experienced can be categorized into three types. (1) Feasible and desirable: The participants found these situations achievable and were interested in trying (e.g., *studying alone at a cafe in the morning* (P05)). They were seen as meaningful strategies that could introduce new ways to reduce stress. (2) Feasible but not desirable: These were situations that, although not impossible to apply, were not appealing to participants (e.g., *taking an online class at the gym in the afternoon* (P04) or *studying with friends at a pub at night* (P10)). These strategies remained as potential options depending on preferences. (3) Infeasible: These combinations were entirely unachievable based on physical limitations, social norms, or common sense (e.g., *studying with my girlfriend in the evening at the dormitory* (P02), since the dormitory is separated by gender).

Although untested situations may likely have been lower in priority when selecting a coping strategy, participants noted that it was still valuable to display them. P04 noted, "*I believe that untried results should still be shown. They're based on my records and might suggest new ways to manage stress. I think it's important to be aware of these possibilities.*" P09 also responded, "*If I only consider situations I've already experienced, I'll never have the chance to try something new, and my stress patterns will probably stay the same.*" Since the What-If feature allowed for simulation before making actual changes, the participants anticipated no risk in exploring a wide range of untried situations.

Number of context changes: The next most important factor for many participants when selecting a coping strategy was the number of context changes. Participants generally preferred strategies that required fewer context changes as they found it less burdensome to achieve the desired outcome. "*To lower my stress, I'd have to switch locations, find someone to do an activity with, and so on... and when all that starts piling up, it just feels more overwhelming for me*" (P09). In addition, they noted that strategies involving more contextual changes became less relevant to the original situation, making it harder to perceive them as practical coping strategies (P01, P03, P05).

The setting of the constraints also influences the number of contexts that can be changed. For instance, participants who fixed both the activity and location in the constraints could identify coping strategies that involved changing social settings and time. In some cases, they did not pay much attention to the number of contexts that could be changed when setting constraints, as P11 noted. "*I took setting constraints to mean that everything except those conditions could be changed. So, how many context changes I allow became less important to me.*"

High-stress probability: High-stress probability was not considered as a primary factor by the participants unlike other factors. This was partly because all the presented coping strategies had a predicted probability of 'high stress' below 50%, which resulted in lower stress than the target situation. Consequently, participants viewed this probability as a minor factor, considering it only when comparing similar coping strategies.

When comparing these probabilities, the participants considered the number of contexts that needed to be changed. They assessed whether the context changes were worthwhile, similar to a cost-benefit analysis. If the probability of reducing stress below a moderate level did not increase significantly with additional context changes (e.g., changing location), they preferred strategies that required fewer changes. P05 noted, "*If changing one context reduces my stress by 20%, and changing two reduces it by 40-50%, I'd probably go for changing two. But if changing two only reduces it by an extra 5%, I don't think I'd bother changing both.*"

Through this evaluation process, participants proposed additional methods to generate and provide more feasible coping strategies. They emphasized that user feedback could play a significant role in refining the strategies. For instance, P04 mentioned, "*It would be great to get user feedback on the strategies shown and decide whether to keep showing them based on that feedback*." Additionally, participants suggested the need for users to set evaluation criteria or prioritize contexts from the beginning. P10 noted that this approach could better account for real-world constraints while introducing untried situations for consideration. "*For example, CounterStress could allow users to prioritize study locations like the library, dorm, or home in advance. This way, they can select realistic places while still adjusting factors like time or social setting*."

5.3 RQ3: How Do Users Apply Coping Strategies Based on Counterfactuals in Real-World Settings?

The one-week field user study specifically focused on understanding how participants applied CounterStress in real-world settings. Insights from diary entries revealed diverse usage behaviors and demonstrated how counterfactual-based coping strategies supported stress management in their daily lives.

5.3.1 Purpose of Using CounterStress and Coping Planning Process. Participants reported using CounterStress for various purposes. They employed it to review stressors based on the collected data and to plan situations for reducing stress before initiating specific activities (e.g., study (P04), club activities (P06), and activities planned for the upcoming semester (P09)). Additionally, some participants used CounterStress in real-time situations to explore ways to lower their stress. Furthermore, they used it at the end of the day to reflect on their stress levels.

The process of identifying an appropriate coping strategy was similar to that used in the lab-based user study. Participants typically fixed the activity contexts that interested them and explored strategies by changing other context types. They also tended to focus on context combinations they had previously experienced to apply the suggested strategies readily. P08 reported, "*I focused on situations I've experienced before and chose ones that (1) significantly lowered the probability and (2) were easier to achieve.*" When using CounterStress in real-life settings, participants considered various criteria to determine the optimal strategy, as P09 noted. "*I set constraints to filter out essential tasks and then looked for better options by balancing the probability change with the number of context changes needed.*" Some participants explored all results without specific constraints, driven by curiosity about the analysis outcomes (P09, P10).

5.3.2 Experience of Utilizing Counterfactual-Based Coping Strategies. The participants reported a range of experiences with using coping strategies in their diaries. Initially, they attempted to change the contexts based on the strategies suggested by CounterStress. P10 reported, "*I adjusted my work hours and started working early in the morning before most of my colleagues arrived.*" When making a new decision, CounterStress serves as a tool to predict the outcomes of these changes before implementation. Some participants experienced a palpable reduction in stress after following these context changes, as P09 mentioned, "*I tried writing my paper at a cafe instead of at home, and with a friend instead of alone. This really helped me keep my stress low while also boosting my productivity.*" Moreover, there were cases in which participants reconfirmed that the provided stress information was consistent with real-world stress. P08 reported, "*I was working alone at night and when a colleague came in, I felt a bit uncomfortable. When I checked CounterStress, I saw that my stress state shifted from low to high when I was with a colleague compared to being alone.*"

The use of CounterStress in daily life also influenced the thoughts and behaviors of participants. They recognized that the impact of context on everyday stress was greater than they had initially assumed. Drawing insights from CounterStress, they aimed to actively apply coping strategies to reduce stress, such as *spending more time in stress-relieving locations* (P06), *completing important tasks early in the day* (P10), and *having meals at a restaurant rather than at the dormitory or convenience store whenever possible* (P08). Some participants found it valuable to try new approaches to stress management that they had not previously considered. Additionally, participants expressed interest in integrating the contextual data with other types of data to gain a more comprehensive understanding of their stress. P09 noted, "*I realized that data from other health apps, like sleep or eating habits, can also affect my stress levels. I'd like to consider this information too and do a more in-depth analysis.*"

In addition to sharing their experiences with CounterStress in daily life, participants identified potential areas for improving the system, particularly by expanding the scope of data and incorporating collective insights from other users. Participants highlighted the importance of collecting more diverse and detailed data to uncover actionable insights. For instance, P08 wrote, "*It would be helpful to collect more detailed context. Like, instead of just 'class,' it could say 'CS101' so the analysis can be more specific to each situation.*" They also proposed integrating data from external sources, such as wearables or journals to enhance the system's recommendations. As P11 noted in their diary, "*Things like exercise and sleep probably affect stress, too. Since I track that info with my smartwatch automatically, I think it'd be great if it could be used in the coping strategies as well.*" Additionally, they suggested leveraging aggregated data from other users to broaden the range of feasible coping strategies. P12 wrote, "*It might be a good idea to show coping strategies that others commonly find feasible. That way, even if it's a situation I haven't experienced, I could still try it out.*" Such aggregated insights could also help exclude impractical strategies, as noted by P11: "*If none of the users have tried a certain situation, it could be excluded.*"

6 Discussion

6.1 Leveraging Counterfactual Explanations for Personal Informatics

In this study, we designed a PI system named CounterStress that supports stress-coping planning using counterfactual explanations. CounterStress is designed to focus on "practical solutions" for stressful situations encountered in daily life, empowering users to make informed decisions through actionable and personalized coping strategies derived from their own data. The system helps users identify necessary changes in specific situations, offering effective ways to reduce stress and improve well-being, even when real-world constraints limit their options. The lab-based and field user studies demonstrated that CounterStress enabled participants to explore diverse coping strategies, providing clear, data-driven insights that helped them manage stress in everyday scenarios.

Previous studies have designed PI systems that analyze and provide insights into factors affecting users' health and well-being [5, 45, 61], or that support the experimentation process [23, 48]. However, critical challenges remained in the coping planning process using personal data, mainly due to multiple, simultaneously occurring stress-influencing factors and practical constraints that limited direct control over stressors. Through our exploration of RQ1, we found that the use of counterfactual explanations in the coping planning process could effectively address these challenges. CounterStress **enabled users to explore coping strategies without the need for complex data analysis.** Furthermore, it generated diverse counterfactual-based coping strategies, each differing in the contextual factors to be changed, **allowing users to select the most suitable options for each situation.** This variety in counterfactuals also **accounted for real-world constraints,** facilitating the provision of feasible and actionable solutions.

Participants could understand contextual factors that *causally* influenced stress through the causal inference analysis provided on the Analysis screen of CounterStress. These causal factors were also considered when exploring coping strategies on the What-If screen. However, when directly addressing these causal factors was not feasible, participants relied on counterfactual explanations to effectively plan alternative coping strategies. Both counterfactual explanations and causal inference aim to answer "*what if?*" questions by analyzing the impact of changing a treatment (i.e., cause) condition on an outcome. While they share this common purpose, they differ fundamentally in methodological rigor and assumptions.

Counterfactual explanations typically analyze data without controlling for confounding variables, relying on correlation-based approaches. In contrast, causal inference ensures that observed

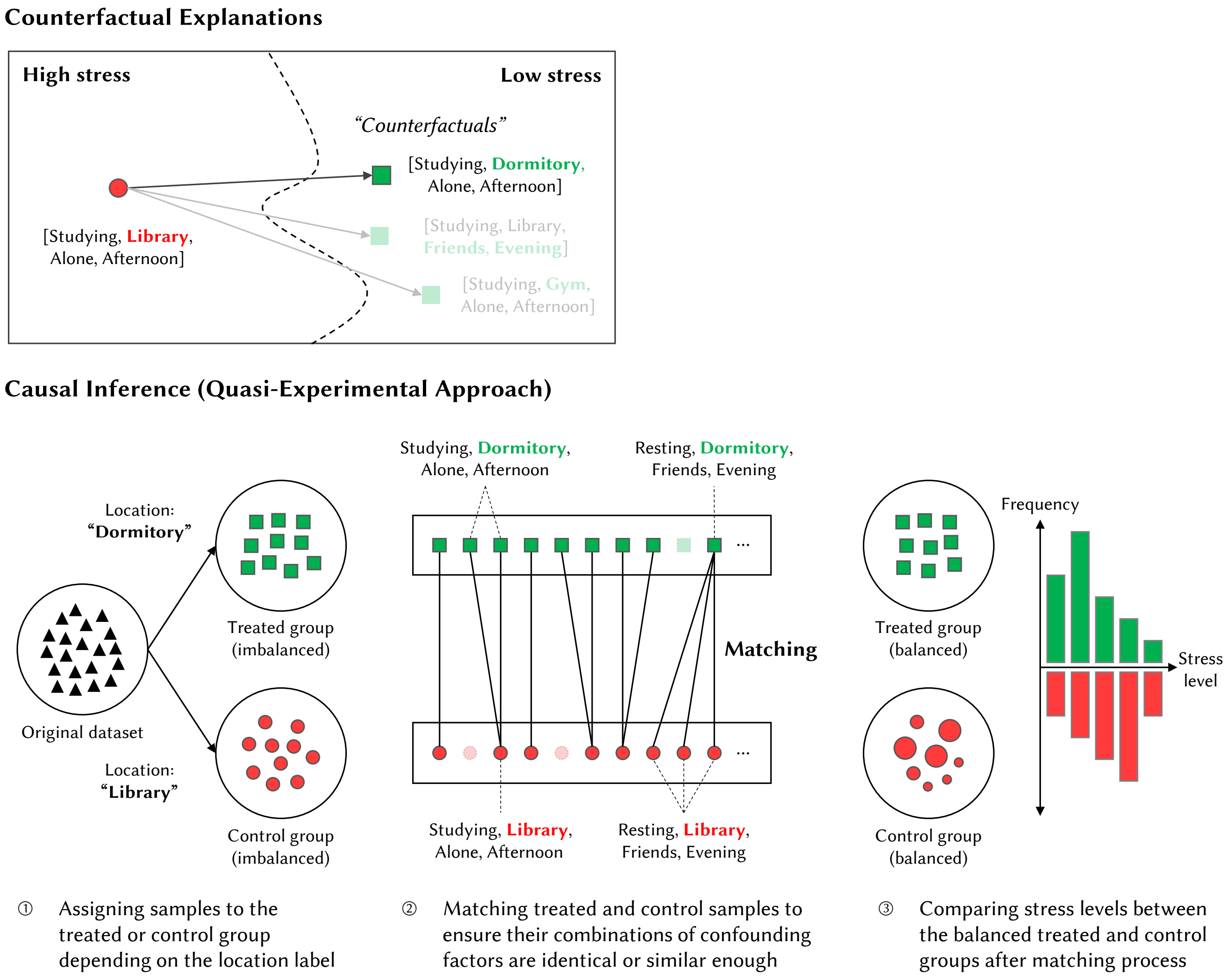


Figure 6: The combination of counterfactual explanations and causal inference: Based on the altered condition in the counterfactual scenario (e.g., from *Library* to *Dormitory*), we can further investigate whether this change is the "cause" of a reduction in stress levels. This is achieved through a quasi-experimental approach, which controls for confounding factors (e.g., activity, social setting, and time at those locations) and estimates causal effects [45].

differences in outcomes are solely attributable to the treatment condition by balancing the distribution of confounding variables across comparison groups. This distinction highlights a limitation of counterfactual explanations: i.e., their lack of control for confounding variables may lead to less rigorous and potentially biased insights. Despite this limitation, exploring counterfactual explanations is still valuable because multiple variables (including both confounders and outcome predictors) can be simultaneously examined to find a feasible alternate reality with a desirable outcome.

To enhance the robustness of findings, as a promising solution, exploring counterfactural explanations can be combined with causal analysis. For instance, Figure 6 illustrates a counterfactual scenario where the location changes from [studying, **library**, alone, afternoon] to [studying, **dormitory**, alone, afternoon]. Here, after viewing this counterfactual scenario, a user may want to subsequently validate a causal relationship. We can apply a quasi-experimental approach, such as matching, as briefly illustrated in Section 3.3; e.g., after setting 'library' as a control group and 'dormitory' as a treated group, remaining contextual factors (activity, social setting, and time) are matched for treatment effect estimation [46]. Consequently, **integrating counterfactual explanations with causal inference enables users to validate the effectiveness of strategies, offering analytical rigor while maintaining exploratory flexibility.** Beyond quasi-experimental approaches, causal ML techniques could also be considered as an alternative for integrating causal estimation into CounterStress [31]. Methods such as causal forests [86] or deep-learning-based treatment effect estimation [64, 78] may provide more scalable and data-driven ways to estimate causal effects. Unlike traditional quasi-experimental methods requiring manual feature selection and predefined model structures, causal ML techniques automatically learn relevant variables and capture complex, nonlinear relationships in high-dimensional data. This enables more flexible and adaptive causal effect estimation while complementing counterfactual explanations.

The process by which CounterStress supports coping planning can be interpreted through the lens of Lazarus and Folkman's transactional model of stress and coping [55]. Based on the insights provided in the Review and Analysis screens, users undergo the *primary appraisal* to identify whether their stress levels are acceptable and which contexts may affect their stress. If users discover that a particular context is a stressor, they then undergo the *secondary appraisal*, evaluating whether they can control or manage that context using their information. During this process, CounterStress provides potential coping strategies derived from counterfactuals. These strategies can be considered either *problem-focused* or *emotion-focused* coping, depending on whether they control the stressor itself or change other contexts to reduce stress. CounterStress can also effectively represent the latter, where stress is managed by indirectly controlling emotional responses instead of directly addressing the stressor. Thus, the coping planning facilitated by CounterStress provides **actionable, context-tailored strategies while aligning closely with established theoretical frameworks,** underscoring its practical relevance and theoretical grounding in stress management.

6.2 Considerations for Generating and Delivering Effective Counterfactuals in Personal Informatics

In this study, we explored how CounterStress could support stress management by generating multiple counterfactuals for target situations. However, findings from the lab-based user study (RQ2) indicated that not all counterfactuals were viable as coping strategies, and users prioritized different criteria when evaluating their applicability. Similarly, insights from the field user study (RQ3) highlighted additional considerations for generating and delivering effective counterfactual-based coping strategies, such as leveraging other users' data and insights. In particular, users gave higher priority to counterfactuals resembling situations they had previously experienced, emphasizing the need to assess their feasibility in real-world contexts. Based on these findings, we propose **approaches to refine the generation and evaluation of counterfactuals to provide users with more feasible and actionable solutions.**

6.2.1 Generating Counterfactuals by Incorporating User Preferences. First, we explored ways to create more relevant coping strategies when generating counterfactuals. Although the overall trends were similar, the users showed slight differences in the criteria they prioritized. These differences can be incorporated into the counterfactual generation process by **assigning weights to the objective functions** in the existing counterfactual explanation methods.

For example, Dandl et al. [22] minimized four objectives (i.e., validity, similarity, minimality, and plausibility) in the loss function. By weighing these objectives based on user priorities, different sets of counterfactuals can be generated. These weights can either be manually set by users or adaptively adjusted based on their feedback on counterfactuals. Consequently, users receive coping strategies that align better with their situation and preferences.

6.2.2 Improving Feasibility of Counterfactuals with Statistical Approaches. In our study, the feasibility was assessed by simply tallying the frequency of a specific situation in the existing dataset. However, this frequency-based approach has limitations since it does not consider the interactions between individual contexts (e.g., "I am often in the *dormitory* at *night*") and fails to provide additional insights for context combinations that have never occurred, as they are simply assigned a frequency of zero. To address this, the likelihood of a counterfactual occurring may be calculated by leveraging **the conditional probabilities between the contexts included in that counterfactual,** thereby using this as a criterion for assessing its feasibility.

For example, if the counterfactual for a target situation [studying, library, alone, afternoon] is [studying, dormitory, alone, afternoon], the probability can be calculated as P(studying, dormitory, alone, afternoon) = P(dormitory | studying, alone, afternoon) · P(studying, alone, afternoon). In addition, by applying smoothing techniques (e.g., Laplace smoothing), a probability of zero no longer needs to be assigned to combinations that have never occurred, thus generating more informative estimates. Furthermore, many situations that are completely unrealistic can be filtered out by providing only the counterfactuals that exceed a certain probability threshold.

6.2.3 Leveraging User Feedback for Feasible Counterfactual Solutions. We propose **employing user feedback on provided counterfactuals** to exclude infeasible counterfactuals from future recommendations as options for a coping strategy. For example, if a counterfactual such as [studying, dormitory, girlfriend, night] is provided, information can be collected from the user regarding subsets of contexts that cannot exist together (e.g., dormitory, girlfriend), whereby counterfactuals containing such subsets are not generated. Alternatively, as suggested in our user study, only combinations that are realistically possible can be considered. However, this approach may remove new possibilities that users have not previously tried. Therefore, it would be preferable to focus on excluding only those combinations that are entirely impossible while supporting users in making the final decision within the remaining counterfactuals.

Additionally, as noted in the user study, common sense can be incorporated as feedback to evaluate the feasibility of counterfactuals. For instance, the general assumption is that people study, work, or hold meetings in libraries, whereas social activities and dining typically occur at bars. Although studying at a bar is not entirely impossible, using common sense, this option would be rarely chosen. Therefore, by incorporating common sense, a lower priority can be assigned to combinations that are unlikely to occur in real life, thereby providing counterfactuals accordingly.

Furthermore, by aggregating feedback from multiple users on various counterfactuals, coping strategies that are generally considered acceptable or preferred can be identified. This approach increases the likelihood of presenting counterfactuals that users are likely to attempt. By continuously updating the criteria for appropriate counterfactuals based on user feedback, the system can offer even more personalized coping strategies.

6.2.4 Delivering Coping Strategies at Opportune Moments. Finally, **opportune moments for providing coping strategies** can be considered to ensure that users can utilize them more effectively [41, 69]. Our field user study revealed that users tended to review these coping strategies before certain situations and utilize them to plan ways to reduce stress. Based on this, we considered the following delivery approaches.

First, by analyzing users' self-tracking data, we can identify their routines and predict stress levels in upcoming situations. If a situation is predicted to be stressful, coping strategies can be provided in advance, allowing users to proactively manage their stress. Additionally, the system may deliver this information when users are reflecting on their day and planning for tomorrow. This would further enhance their stress management skills and capabilities by leveraging their self-knowledge from PI systems.

6.3 Limitations and Future Work

Our user study results indicated that counterfactual-based coping strategies have the potential to be effective in managing daily stress. A more rigorous evaluation of CounterStress requires conducting controlled experiments with larger samples. The system's effectiveness can be assessed by comparing the number of coping strategies generated by users after using the system and the actual reduction in stress after applying these strategies. Additionally, continuous data collection is essential for determining whether newly generated strategies remain effective over time.

In our study, the entire data consisted of categorical variables, which may have limited the generation of counterfactuals. For instance, the number of combinations may be limited, and subtle changes may be more difficult to reflect than numerical variables. Therefore, future research can consider generating counterfactuals by incorporating numerical variables such as step count or sleep duration, which can influence stress levels. Utilizing passive sensor data collected from smartphones or wearables can be an option for enabling this, with the added benefit of reducing the burden of data collection on users. Incorporating more detailed and fine-grained contextual information can further enhance the effectiveness of counterfactual-based strategies.

Finally, we suggest exploring future research beyond stress coping because the proposed approach is not limited to stress-related scenarios and can be adapted to other types of data. For example, when monitoring blood glucose levels in diabetic patients, counterfactual-based strategies can generate personalized coping strategies for problematic situations. Key factors like sleep quality, physical activity, and carbohydrate intake can be tracked, and strategies using these factors can be delivered when glucose levels are expected to exceed acceptable ranges. These strategies may be pre-simulated or provided as concrete actions when glucose levels increase. Similarly, this counterfactual-based approach can be extended to other health and behavioral scenarios, offering tailored solutions targeting specific situations.

7 Conclusion

In this study, we introduced a PI system, *CounterStress,* which leverages counterfactual explanations to help users plan coping strategies for high-stress situations. We investigated the user experience with CounterStress to understand how PI users explore and apply counterfactual-based coping strategies. Our study demonstrated the feasibility of this approach for stress-coping planning, presenting how users evaluate and prioritize the different counterfactuals suggested by the system. We also uncovered several design implications for generating and delivering counterfactual-based solutions that can be more personalized and effective for stress management. Additionally, we found that counterfactual explanations have the potential to expand users' self-awareness and improve their ability to manage health-related challenges. By integrating this approach with existing PI systems, we believe that our system can offer better support not only for self-reflection but also for providing practical guidance leading to meaningful actions.

Acknowledgments

We thank Jiwon Jung and the members of the Interactive Computing Lab for their active discussions and insightful feedback on the design of CounterStress. This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Korean government (MSIT) (2022R1A2C2011536).

References

[1] Phil Adams, Mashfiqui Rabbi, Tauhidur Rahman, Mark Matthews, Amy Voida, Geri Gay, Tanzeem Choudhury, and Stephen Voida. 2014. Towards Personal Stress Informatics: Comparing Minimally Invasive Techniques for Measuring Daily Stress in the Wild. In *Proceedings of the 8th International Conference on Pervasive Computing Technologies for Healthcare.* ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), Brussels, BEL, 72–79. doi:10.4108/icst.pervasivehealth.2014.254959

[2] Mona Alhasani, Dinesh Mulchandani, Oladapo Oyebode, Nilufar Baghaei, and Rita Orji. 2022. A Systematic and Comparative Review of Behavior Change Strategies in Stress Management Apps: Opportunities for Improvement. *Frontiers in Public Health* 10 (2022), 777567. doi:10.3389/fpubh.2022.777567

[3] J.M. Christian Bastien. 2010. Usability Testing: A Review of Some Methodological and Technical Aspects of the Method. *International Journal of Medical Informatics* 79, 4 (2010), e18–e23. doi:10.1016/j.ijmedinf.2008.12.004

[4] Eric PS Baumer, Vera Khovanskaya, Mark Matthews, Lindsay Reynolds, Victoria Schwanda Sosik, and Geri Gay. 2014. Reviewing Reflection: On the Use of Reflection in Interactive System Design. In *Proceedings of the 2014 Conference on Designing Interactive Systems.* Association for Computing Machinery, New York, NY, USA, 93–102. doi:10.1145/2598510.2598598

[5] Frank Bentley, Konrad Tollmar, Peter Stephenson, Laura Levy, Brian Jones, Scott Robertson, Ed Price, Richard Catrambone, and Jeff Wilson. 2013. Health Mashups: Presenting Statistical Patterns between Wellbeing Data and Context in Natural Language to Promote Behavior Change. *ACM Transactions on Computer-Human Interaction (TOCHI)* 20, 5 (2013), 1–27. doi:10.1145/2503823

[6] Marit Bentvelzen, Paweł W Woźniak, Pia SF Herbes, Evropi Stefanidi, and Jasmin Niess. 2022. Revisiting Reflection in Hci: Four Design Resources for Technologies That Support Reflection. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 1 (2022), 1–27. doi:10.1145/3517233

[7] Matthew Blackwell, Stefano Iacus, Gary King, and Giuseppe Porro. 2009. cem: Coarsened Exact Matching in Stata. *The Stata Journal* 9, 4 (2009), 524–546. doi:10.1177/1536867X0900900402

[8] Andrey Bogomolov, Bruno Lepri, Michela Ferron, Fabio Pianesi, and Alex Pentland. 2014. Daily Stress Recognition from Mobile Phone Data, Weather Conditions and Individual Traits. In *Proceedings of the 22nd ACM international conference on Multimedia.* Association for Computing Machinery, New York, NY, USA, 477–486. doi:10.1145/2647868.2654933

[9] Niall Bolger, Angelina Davis, and Eshkol Rafaeli. 2003. Diary Methods: Capturing Life as it is Lived. *Annual Review of Psychology* 54 (2003), 579–616. doi:10.1146/annurev.psych.54.101601.145030

[10] Virginia Braun and Victoria Clarke. 2006. Using Thematic Analysis in Psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101. doi:10.1191/1478088706QP063OA
[11] John Brooke. 1996. SUS: A 'Quick and Dirty' Usability. *Usability Evaluation in Industry* 189, 3 (1996), 189–194.
[12] Ruth MJ Byrne. 2016. Counterfactual Thought. *Annual Review of Psychology* 67, 1 (2016), 135–157. doi:10.1146/annurev-psych-122414-033249
[13] Yekta Said Can, Bert Arnrich, and Cem Ersoy. 2019. Stress Detection in Daily Life Scenarios Using Smart Phones and Wearable Sensors: A Survey. *Journal of Biomedical Informatics* 92 (2019), 103139. doi:10.1016/j.jbi.2019.103139
[14] Scott Carter and Jennifer Mankoff. 2005. When Participants Do the Capturing: The Role of Media in Diary Studies. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 899–908. doi:10.1145/1054972.1055098
[15] Janghee Cho, Tian Xu, Abigail Zimmermann-Niefield, and Stephen Voida. 2022. Reflection in Theory and Reflection in Practice: An Exploration of the Gaps in Reflection Support among Personal Informatics Apps. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–23. doi:10.1145/3491102.3501991
[16] Eun Kyoung Choe, Bongshin Lee, and M.C. Schraefel. 2015. Characterizing Visualization Insights from Quantified Selfers' Personal Data Presentations. *IEEE Computer Graphics and Applications* 35, 4 (2015), 28–37. doi:10.1109/MCG.2015.51
[17] Eun Kyoung Choe, Bongshin Lee, Haining Zhu, Nathalie Henry Riche, and Dominikus Baur. 2017. Understanding Self-Reflection: How People Reflect on Personal Data Through Visual Data Exploration. In *Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare*. Association for Computing Machinery, New York, NY, USA, 173–182. doi:10.1145/3154862.3154881
[18] Eun Kyoung Choe, Nicole B Lee, Bongshin Lee, Wanda Pratt, and Julie A Kientz. 2014. Understanding Quantified-Selfers' Practices in Collecting and Exploring Personal Data. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1143–1152. doi:10.1145/2556288.2557372
[19] George P Chrousos and Philip W Gold. 1992. The Concepts of Stress and Stress System Disorders: Overview of Physical and Behavioral Homeostasis. *Jama* 267, 9 (1992), 1244–1252. doi:10.1001/jama.1992.03480090092034
[20] Sheldon Cohen, Tom Kamarck, and Robin Mermelstein. 1994. Perceived Stress Scale. *Measuring Stress: A Guide for Health and Social Scientists* 10, 2 (1994), 1–2.
[21] Sandra M. Coulon, Courtney M. Monroe, and Delia S. West. 2016. A Systematic, Multi-domain Review of Mobile Smartphone Apps for Evidence-Based Stress Management. *American Journal of Preventive Medicine* 51, 1 (2016), 95–105. doi:10.1016/j.amepre.2016.01.026
[22] Susanne Dandl, Christoph Molnar, Martin Binder, and Bernd Bischl. 2020. Multi-Objective Counterfactual Explanations. In *Parallel Problem Solving from Nature – PPSN XVI*. Springer International Publishing, Cham, 448–469. doi:10.1007/978-3-030-58112-1_31
[23] Nediyana Daskalova, Danaë Metaxa-Kakavouli, Adrienne Tran, Nicole Nugent, Julie Boergers, John McGeary, and Jeff Huang. 2016. SleepCoacher: A Personalized Automated Self-Experimentation System for Sleep Recommendations. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*. Association for Computing Machinery, New York, NY, USA, 347–358. doi:10.1145/2984511.2984534
[24] Lianne P De Vries, Bart ML Baselmans, and Meike Bartels. 2021. Smartphone-Based Ecological Momentary Assessment of Well-Being: A Systematic Review and Recommendations for Future Studies. *Journal of Happiness Studies* 22, 5 (2021), 2361–2408. doi:10.1007/s10902-020-00324-7
[25] Amit Dhurandhar, Pin-Yu Chen, Ronny Luss, Chun-Chen Tu, Paishun Ting, Karthikeyan Shanmugam, and Payel Das. 2018. Explanations Based on the Missing: Towards Contrastive Explanations With Pertinent Negatives. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, Vol. 31. Curran Associates Inc., Red Hook, NY, USA, 590–601. doi:10.5555/3326943.3326998
[26] Xianghua (Sharon) Ding, Shuhan Wei, Xinning Gui, Ning Gu, and Peng Zhang. 2021. Data Engagement Reconsidered: A Study of Automatic Stress Tracking Technology in Use. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3411764.3445763
[27] Rudresh Dwivedi, Devam Dave, Het Naik, Smiti Singhal, Rana Omer, Pankesh Patel, Bin Qian, Zhenyu Wen, Tejal Shah, Graham Morgan, et al. 2023. Explainable AI (XAI): Core Ideas, Techniques, and Solutions. *Comput. Surveys* 55, 9 (2023), 1–33. doi:10.1145/3561048
[28] Daniel A. Epstein, Clara Caldeira, Mayara Costa Figueiredo, Xi Lu, Lucas M. Silva, Lucretia Williams, Jong Ho Lee, Qingyang Li, Simran Ahuja, Qiuer Chen, Payam Dowlatyari, Craig Hilby, Sazeda Sultana, Elizabeth V. Eikey, and Yunan Chen. 2020. Mapping and Taking Stock of the Personal Informatics Literature. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 4 (2020), 1–38. doi:10.1145/3432231
[29] Daniel A. Epstein, Nicole B. Lee, Jennifer H. Kang, Elena Agapie, Jessica Schroeder, Laura R. Pina, James Fogarty, Julie A. Kientz, and Sean Munson. 2017. Examining Menstrual Tracking to Inform the Design of Personal Informatics Tools. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 6876–6888. doi:10.1145/3025453.3025635
[30] Daniel A Epstein, An Ping, James Fogarty, and Sean A Munson. 2015. A Lived Informatics Model of Personal Informatics. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. Association for Computing Machinery, New York, NY, USA, 731–742. doi:10.1145/2750858.2804250
[31] Stefan Feuerriegel, Dennis Frauen, Valentyn Melnychuk, Jonas Schweisthal, Konstantin Hess, Alicia Curth, Stefan Bauer, Niki Kilbertus, Isaac S Kohane, and Mihaela van der Schaar. 2024. Causal Machine Learning for Predicting Treatment Outcomes. *Nature Medicine* 30, 4 (2024), 958–968. doi:10.1038/s41591-024-02902-1
[32] Rowanne Fleck and Geraldine Fitzpatrick. 2010. Reflecting on Reflection: Framing a Design Landscape. In *Proceedings of the 22nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction*. Association for Computing Machinery, New York, NY, USA, 216–223. doi:10.1145/1952222.1952269
[33] Susan Folkman. 2012. Stress, Coping, and Hope. In *Psychological Aspects of Cancer*. Springer, Boston, MA, USA, 119–127. doi:10.1007/978-1-4614-4866-2_8
[34] Susan Folkman and Judith Tedlie Moskowitz. 2004. Coping: Pitfalls and Promise. *Annu. Rev. Psychol.* 55, 1 (2004), 745–774. doi:10.1146/annurev.psych.55.090902.141456
[35] Susannah Fox and Maeve Duggan. 2013. Tracking for Health.
[36] Jerome H Friedman. 2001. Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics* 29, 5 (2001), 1189–1232. doi:10.1214/aos/1013203451
[37] Giorgos Giannakakis, Dimitris Grigoriadis, Katerina Giannakaki, Olympia Simantiraki, Alexandros Roniotis, and Manolis Tsiknakis. 2022. Review on Psychological Stress Detection Using Biosignals. *IEEE Transactions on Affective Computing* 13, 1 (2022), 440–460. doi:10.1109/TAFFC.2019.2927337
[38] Randy Goebel, Ajay Chander, Katharina Holzinger, Freddy Lecue, Zeynep Akata, Simone Stumpf, Peter Kieseberg, and Andreas Holzinger. 2018. Explainable AI: the New 42?. In *Machine Learning and Knowledge Extraction*. Springer International Publishing, Cham, 295–303. doi:10.1007/978-3-319-99740-7_21
[39] Alex Goldstein, Adam Kapelner, Justin Bleich, and Emil Pitkin. 2015. Peeking Inside the Black Box: Visualizing Statistical Learning With Plots of Individual Conditional Expectation. *Journal of Computational and Graphical Statistics* 24, 1 (2015), 44–65. doi:10.1080/10618600.2014.907095
[40] Riccardo Guidotti. 2022. Counterfactual Explanations and How to Find Them: Literature Review and Benchmarking. *Data Mining and Knowledge Discovery* 38 (2022), 1–55. doi:10.1007/s10618-022-00831-6
[41] Wendy Hardeman, Julie Houghton, Kathleen Lane, Andy Jones, and Felix Naughton. 2019. A Systematic Review of Just-In-Time Adaptive Interventions (JITAIs) to Promote Physical Activity. *International Journal of Behavioral Nutrition and Physical Activity* 16 (2019), 1–21. doi:10.1186/s12966-019-0792-7
[42] Karen Hovsepian, Mustafa Al'Absi, Emre Ertin, Thomas Kamarck, Motohiro Nakajima, and Santosh Kumar. 2015. cStress: Towards a Gold Standard for Continuous Stress Assessment in the Mobile Environment. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. Association for Computing Machinery, New York, NY, USA, 493–504. doi:10.1145/2750858.2807526
[43] Esther Howe, Jina Suh, Mehrab Bin Morshed, Daniel McDuff, Kael Rowan, Javier Hernandez, Marah Ihab Abdin, Gonzalo Ramos, Tracy Tran, and Mary P Czerwinski. 2022. Design of Digital Workplace Stress-Reduction Intervention Systems: Effects of Intervention Type and Timing. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–16. doi:10.1145/3491102.3502027
[44] Stefano M Iacus, Gary King, and Giuseppe Porro. 2012. Causal Inference without Balance Checking: Coarsened Exact Matching. *Political Analysis* 20, 1 (2012), 1–24. doi:10.1093/pan/mpr013
[45] Gyuwon Jung, Sangjun Park, and Uichin Lee. 2024. DeepStress: Supporting Stressful Context Sensemaking in Personal Informatics Systems Using a Quasi-experimental Approach. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–18. doi:10.1145/3613904.3642766
[46] Gyuwon Jung, Sangjun Park, Eun-Yeol Ma, Heeyoung Kim, and Uichin Lee. 2024. A Tutorial on Matching-based Causal Analysis of Human Behaviors Using Smartphone Sensor Data. *Comput. Surveys* 56, 9 (2024), 1–33. doi:10.1145/3648356
[47] Soowon Kang, Woohyeok Choi, Cheul Young Park, Narae Cha, Auk Kim, Ahsan Habib Khandoker, Leontios Hadjileontiadis, Heepyung Kim, Yong Jeong, and Uichin Lee. 2023. K-EmoPhone: A Mobile and Wearable Dataset with In-Situ Emotion, Stress, and Attention Labels. *Scientific Data* 10, 1 (2023), 351. doi:10.1038/s41597-023-02248-2
[48] Ravi Karkar, Jessica Schroeder, Daniel A Epstein, Laura R Pina, Jeffrey Scofield, James Fogarty, Julie A Kientz, Sean A Munson, Roger Vilardaga, and Jasmine

Zia. 2017. TummyTrials: A Feasibility Study of Using Self-Experimentation to Detect Individualized Food Triggers. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 6850–6863. doi:10.1145/3025453.3025480

[49] Elisabeth T Kersten-van Dijk, Joyce HDM Westerink, Femke Beute, and Wijnand A IJsselsteijn. 2017. Personal Informatics, Self-Insight, and Behavior Change: A Critical Review of Current Literature. *Human–Computer Interaction* 32, 5-6 (2017), 268–296. doi:10.1080/07370024.2016.1276456

[50] Inyeop Kim, Hwarang Goh, Nematjon Narziev, Youngtae Noh, and Uichin Lee. 2020. Understanding User Contexts and Coping Strategies for Context-aware Phone Distraction Management System Design. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 4 (2020), 1–33. doi:10.1145/3432213

[51] Taewan Kim, Haesoo Kim, Ha Yeon Lee, Hwarang Goh, Shakhboz Abdigapporov, Mingon Jeong, Hyunsung Cho, Kyungsik Han, Youngtae Noh, Sung-Ju Lee, and Hwajung Hong. 2022. Prediction for Retrospection: Integrating Algorithmic Stress Prediction Into Personal Informatics Systems for College Students' Mental Health. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–20. doi:10.1145/3491102.3517701

[52] Rafal Kocielnik and Natalia Sidorova. 2015. Personalized Stress Management: Enabling Stress Monitoring With LifelogExplorer. *KI-Künstliche Intelligenz* 29 (2015), 115–122. doi:10.1007/s13218-015-0348-1

[53] Rafal Kocielnik, Lillian Xiao, Daniel Avrahami, and Gary Hsieh. 2018. Reflection Companion: A Conversational System for Engaging Users in Reflection on Physical Activity. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 2 (2018), 1–26. doi:10.1145/3214273

[54] Reed Larson and Mihaly Csikszentmihalyi. 2014. *The Experience Sampling Method*. Springer Netherlands, Dordrecht, 21–34. doi:10.1007/978-94-017-9088-8_2

[55] Richard S Lazarus. 1984. *Stress, Appraisal, and Coping*. Vol. 464. Springer, Boston, MA, USA.

[56] Richard S Lazarus. 1986. Puzzles in the Study of Daily Hassles. In *Development as Action in Context: Problem Behavior and Normal Youth Development*. Springer Berlin Heidelberg, Berlin, Heidelberg, 39–53. doi:10.1007/978-3-662-02475-1_3

[57] Richard S Lazarus. 1993. Coping Theory and Research: Past, Present, and Future. *Psychosomatic Medicine* 55, 3 (1993), 234–247. doi:10.1097/00006842-199305000-00002

[58] Kwangyoung Lee, Hyewon Cho, Kobiljon Toshnazarov, Nematjon Narziev, So Young Rhim, Kyungsik Han, YoungTae Noh, and Hwajung Hong. 2020. Toward Future-Centric Personal Informatics: Expecting Stressful Events and Preparing Personalized Interventions in Stress Management. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3313831.3376475

[59] Ian Li, Anind Dey, and Jodi Forlizzi. 2010. A Stage-Based Model of Personal Informatics Systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 557–566. doi:10.1145/1753326.1753409

[60] Ian Li, Anind K Dey, and Jodi Forlizzi. 2011. Understanding My Data, Myself: Supporting Self-Reflection With Ubicomp Technologies. In *Proceedings of the 13th International Conference on Ubiquitous Computing*. Association for Computing Machinery, New York, NY, USA, 405–414. doi:10.1145/2030112.2030166

[61] Zilu Liang, Bernd Ploderer, Wanyu Liu, Yukiko Nagata, James Bailey, Lars Kulik, and Yuxuan Li. 2016. SleepExplorer: A Visualization Tool to Make Sense of Correlations Between Personal Sleep Data and Contextual Factors. *Personal and Ubiquitous Computing* 20 (2016), 985–1000. doi:10.1007/s00779-016-0960-6

[62] Scott M. Lundberg and Su-In Lee. 2017. A Unified Approach to Interpreting Model Predictions. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*. Curran Associates Inc., Red Hook, NY, USA, 4768–4777. doi:10.5555/3295222.3295230

[63] Yuhan Luo, Peiyi Liu, and Eun Kyoung Choe. 2019. Co-Designing Food Trackers with Dietitians: Identifying Design Opportunities for Food Tracker Customization. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3290605.3300822

[64] Yunan Luo, Jian Peng, and Jianzhu Ma. 2020. When Causal Inference Meets Deep Learning. *Nature Machine Intelligence* 2, 8 (2020), 426–427. doi:10.1038/s42256-020-0218-x

[65] Deborah Lupton. 2017. Introduction: Self-Tracking, Health and Medicine. In *Self-Tracking, Health and Medicine*. Routledge, London, UK, 1–5.

[66] Abhinav Mehrotra, Fani Tsapeli, Robert Hendley, and Mirco Musolesi. 2017. MyTraces: Investigating Correlation and Causation Between Users' Emotional States and Mobile Phone Interaction. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 1–21. doi:10.1145/3130948

[67] David C Mohr, Mi Zhang, and Stephen M Schueller. 2017. Personal Sensing: Understanding Mental Health Using Ubiquitous Sensors and Machine Learning. *Annual Review of Clinical Psychology* 13 (2017), 23–47. doi:10.1146/annurev-clinpsy-032816-044949

[68] Ramaravind K Mothilal, Amit Sharma, and Chenhao Tan. 2020. Explaining Machine Learning Classifiers Through Diverse Counterfactual Explanations. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. Association for Computing Machinery, New York, NY, USA, 607–617. doi:10.1145/3351095.3372850

[69] Inbal Nahum-Shani, Shawna N Smith, Bonnie J Spring, Linda M Collins, Katie Witkiewitz, Ambuj Tewari, and Susan A Murphy. 2018. Just-In-Time Adaptive Interventions (JITAIs) in Mobile Health: Key Components and Design Principles for Ongoing Health Behavior Support. *Annals of Behavioral Medicine* 52, 6 (2018), 446–462. doi:10.1007/s12160-016-9830-8

[70] Sameer Neupane, Mithun Saha, Nasir Ali, Timothy Hnat, Shahin Alan Samiei, Anandatirtha Nandugudi, David M Almeida, and Santosh Kumar. 2024. Momentary Stressor Logging and Reflective Visualizations: Implications for Stress Management with Wearables. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–19. doi:10.1145/3613904.3642662

[71] Rafael Poyiadzi, Kacper Sokol, Raul Santos-Rodriguez, Tijl De Bie, and Peter Flach. 2020. FACE: Feasible and Actionable Counterfactual Explanations. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*. Association for Computing Machinery, New York, NY, USA, 344–350. doi:10.1145/3375627.3375850

[72] Nora Ptakauskaite, Anna L Cox, and Nadia Berthouze. 2018. Knowing What You're Doing or Knowing What to Do: How Stress Management Apps Support Reflection and Behaviour Change. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–6. doi:10.1145/3170427.3188648

[73] Amon Rapp, Alessandro Marcengo, Luca Buriano, Giancarlo Ruffo, Mirko Lai, and Federica Cena. 2018. Designing a Personal Informatics System for Users without Experience in Self-Tracking: A Case Study. *Behaviour & Information Technology* 37, 4 (2018), 335–366. doi:10.1080/0144929X.2018.1436592

[74] Amon Rapp and Maurizio Tirassa. 2017. Know Thyself: A Theory of the Self for Personal Informatics. *Human–Computer Interaction* 32, 5-6 (2017), 335–380. doi:10.1080/07370024.2017.1285704

[75] Tabea Reuter and Ralf Schwarzer. 2012. *Manage Stress at Work Through Preventive and Proactive Coping*. John Wiley & Sons, Ltd, Hoboken, NJ, USA, Chapter 27, 499–515. doi:10.1002/9781119206422.ch27

[76] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. "Why Should I Trust You?" Explaining the Predictions of Any Classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Association for Computing Machinery, New York, NY, USA, 1135–1144. doi:10.1145/2939672.2939778

[77] John Rooksby, Mattias Rost, Alistair Morrison, and Matthew Chalmers. 2014. Personal Tracking as Lived Informatics. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1163–1172. doi:10.1145/2556288.2557039

[78] Patrick Schwab, Lorenz Linhardt, Stefan Bauer, Joachim M. Buhmann, and Walter Karlen. 2020. Learning Counterfactual Representations for Estimating Individual Dose-Response Curves. *Proceedings of the AAAI Conference on Artificial Intelligence* 34, 04 (2020), 5612–5619. doi:10.1609/aaai.v34i04.6014

[79] Ralf Schwarzer and Ute Schulz. 2003. *Stressful Life Events*. John Wiley & Sons, Ltd, Chichester, West Sussex, UK, Chapter 2, 25–49. doi:10.1002/0471264385.wei0902

[80] Hans Selye. 1976. *The Stress of Life, Rev.* McGraw Hill, New York, NY, USA.

[81] Moushumi Sharmin, Andrew Raij, David Epstien, Inbal Nahum-Shani, J Gayle Beck, Sudip Vhaduri, Kenzie Preston, and Santosh Kumar. 2015. Visualization of Time-Series Sensor Data to Inform the Design of Just-in-Time Adaptive Stress Interventions. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. Association for Computing Machinery, New York, NY, USA, 505–516. doi:10.1145/2750858.2807537

[82] Xin Tong, Matthew Louis Mauriello, Marco Antonio Mora-Mendoza, Nina Prabhu, Jane Paik Kim, and Pablo E Paredes Castro. 2023. Just Do Something: Comparing Self-proposed and Machine-recommended Stress Interventions among Online Workers with Home Sweet Office. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–20. doi:10.1145/3544548.3581319

[83] Niels Van Berkel, Denzil Ferreira, and Vassilis Kostakos. 2017. The Experience Sampling Method on Mobile Devices. *ACM Computing Surveys (CSUR)* 50, 6 (2017), 1–40. doi:10.1145/3123988

[84] Sahil Verma, John Dickerson, and Keegan Hines. 2020. Counterfactual Explanations for Machine Learning: A Review. *arXiv Preprint arXiv:2010.10596* 2 (2020), 1. doi:10.48550/arXiv.2010.10596

[85] Sandra Wachter, Brent Mittelstadt, and Chris Russell. 2017. Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR. *Harv. JL & Tech.* 31 (2017), 841. doi:10.2139/ssrn.3063289

[86] Stefan Wager and Susan Athey. 2018. Estimation and Inference of Heterogeneous Treatment Effects using Random Forests. *J. Amer. Statist. Assoc.* 113, 523 (2018), 1228–1242. doi:10.1080/01621459.2017.1319839

[87] Rui Wang, Fanglin Chen, Zhenyu Chen, Tianxing Li, Gabriella Harari, Stefanie Tignor, Xia Zhou, Dror Ben-Zeev, and Andrew T Campbell. 2014. StudentLife: Assessing Mental Health, Academic Performance and Behavioral Trends of College

Students Using Smartphones. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. Association for Computing Machinery, New York, NY, USA, 3–14. doi:10.1145/2632048.2632054

[88] Xinghui (Erica) Yan, Loubna Baroudi, Rongqi Bei, Leila Boudalia, Stephen M Cain, Kira Barton, K. Alex Shorter, and Mark W. Newman. 2024. Proposing a Context-informed Layer-based Framework: Incorporating Context into Designing mHealth Technology for Fatigue Management. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference*. Association for Computing Machinery, New York, NY, USA, 571–583. doi:10.1145/3643834.3661615

[89] Panyu Zhang, Gyuwon Jung, Jumabek Alikhanov, Uzair Ahmed, and Uichin Lee. 2024. A Reproducible Stress Prediction Pipeline with Mobile Sensor Data. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 8, 3 (2024), 1–35. doi:10.1145/3678578