

Original Paper

Causal AI Recommendation System for Digital Mental Health: Bayesian Decision-Theoretic Analysis

Mathew Varidel¹, PhD; Victor An¹; Ian B Hickie¹, MD; Sally Cripps², PhD; Roman Marchant², PhD; Jan Scott³, MD; Jacob J Crouse¹, PhD; Adam Poulsen¹, PhD; Bridianne O'Dea⁴, PhD; Sarah McKenna¹, PhD; Frank Iorfino¹, PhD

¹Brain and Mind Centre, The University of Sydney, Sydney, Australia

²Human Technology Institute, University of Technology Sydney, Sydney, Australia

³Institute of Neuroscience, Academic Psychiatry, Newcastle University, Newcastle, United Kingdom

⁴Institute for Mental Health and Wellbeing, Flinders University, Adelaide, Australia

Corresponding Author:

Mathew Varidel, PhD
Brain and Mind Centre
The University of Sydney
94 Mallett Street
Sydney, 2050
Australia
Phone: 61 0293510774
Email: mathew.varidel@sydney.edu.au

Abstract

Background: Digital mental health tools promise to enhance the reach and quality of care. Current tools often recommend content to individuals, typically using generic knowledge-based systems or predictive artificial intelligence (AI). However, predictive AI is problematic for intervention recommendations as cause-effect relationships can be confounded in observed data. Therefore, causal AI is required to compare future outcomes under different interventions.

Objective: We aimed to develop a causal AI recommendation system that uses an individual's current presentation, their preferences, and the learned dynamics between domains to rank interventions.

Methods: We frame the recommendation problem within a Bayesian decision-theoretic framework, whereby a preference ordering of decisions is estimated using the expected utility of outcomes under interventions. The causal processes are assumed to follow a structural causal model, where the posterior distribution of structural causal models is estimated using a Markov chain Monte Carlo method. Expected utilities under interventions are estimated using a do-operation, which estimates the effects of changing a variable on outcomes, while accounting for confounders. We apply our approach to rank domains relating to mental health and well-being as intervention targets for adults (n=619) who used the Innowell Fitness app between September 2021 to September 2023 and completed a questionnaire at 2 time points (1 wk-6 mo from baseline).

Results: The causal AI recommendation system recommends intervention targets as a function of a user's baseline presentation, the causal effects of the intervention on itself and other domains, and the utility function. In our example, psychological distress was typically the optimal intervention target in complex cases where multiple domains were unhealthy at baseline, due to it affecting multiple domains with paths to personal functioning (probability [p] of path; $p_{\text{path}}=86\%$), social support ($p_{\text{path}}=92\%$), sleep ($p_{\text{path}}=88\%$), and physical activity ($p_{\text{path}}=86\%$). The probability of being the optimal intervention target was personal functioning ($p_{\text{opt}}=30\%$), psychological distress ($p_{\text{opt}}=29\%$), social support ($p_{\text{opt}}=18\%$), nutrition ($p_{\text{opt}}=9.6\%$), substance use ($p_{\text{opt}}=6.7\%$), sleep ($p_{\text{opt}}=4.5\%$), and physical activity ($p_{\text{opt}}=2.2\%$).

Conclusions: This work illustrates the use of causality and decision-theoretic principles to personalize interventions in digital mental health tools.

J Med Internet Res 2025;27:e71305; doi: [10.2196/71305](https://doi.org/10.2196/71305)

Keywords: causal inference; causal structure learning; decision theory; decision analysis; digital health; well-being; psychological distress; functioning; sleep; social support

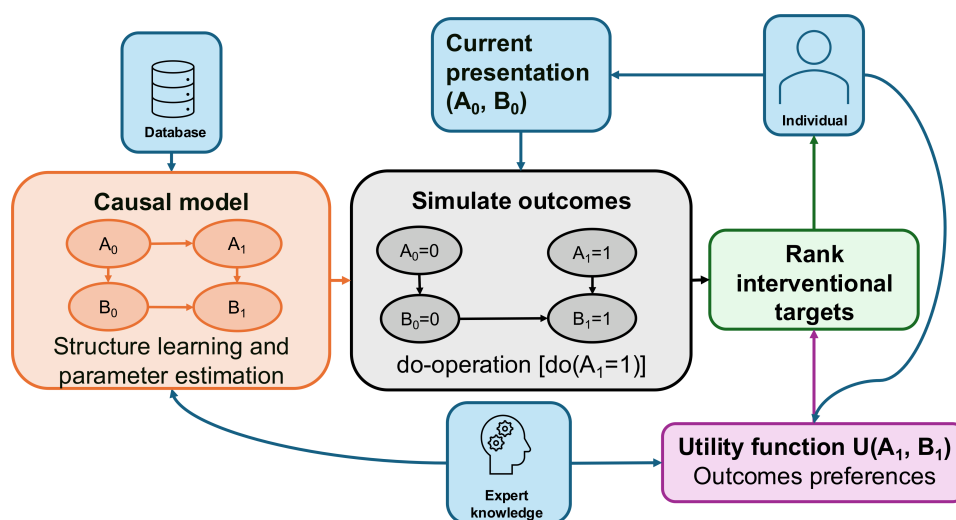
Theory suggests that mental ill-health and poor well-being are emergent phenomena from the interactions between symptoms within and across domains over time [1-4]. This view recognizes that domains—clinical symptoms (eg, depressive and anxiety symptoms), physical health and activity, social support, personal functioning, sleep, nutrition, alcohol or other substance use—influence each other in complex ways. Furthermore, mental ill-health and well-being are influenced by cultural, socioeconomic, or population-level factors [5,6]. This provides a landscape of potential interventional targets that could alleviate mental ill-health and improve well-being. However, this landscape raises the problem of deciding between interventions, which is made difficult by (1) the heterogeneous presentation of mental ill-health and poor well-being, (2) uncertainty about outcomes under different interventions, (3) timeliness, duration, and sequence of interventions, (4) comparison of multidimensional outcomes given individual-level differences in the prioritization of different symptoms or domains, and (5) costs (monetary or otherwise) associated with interventions.

Recommendation systems (RSs) are algorithms that filter content or actions for end users. RSs are often built to filter content to align with an individual’s prior choices or elicited

This approach requires methods to predict outcomes under interventions. The use of predictive AI algorithms to align recommendations with outcomes can be problematic as it estimates relationships between inputs and outcomes using potentially confounded conditional probabilities inferred from observational data [15]. As an alternative, we will explore causal AI [16], which includes mechanisms to estimate causal effects from observational data.

Many digital technologies aimed at improving mental health and well-being exist today, but their reliance on resemblance, knowledge-based, or predictive AI algorithms limits their ability to provide personalized interventional recommendations [17-19]. While there has been interest in building causal AI applications for decision-support systems within health care generally [20,21], there has been very little work on causal AI applications to digital mental health care. Furthermore, few applications of causal AI have been deployed in digital health in general, as it is a relatively new field that requires further understanding and tools to find real-world applicability. We will outline a causal artificial intelligence recommendation system (CAIRS) that aims to improve interventional decision-making in digital mental health applications (Figure 1). We will then show an application to a general population cohort to rank domains as intervention targets in the mental health and well-being context.

Figure 1. CAIRS uses expert knowledge and observational data to learn a causal model, which is then used to simulate outcomes under idealized interventions and displays intervention targets based on their expected utility. CAIRS: causal artificial intelligence recommendation system.



Methods

Ethical Considerations

This study analyzed nonidentifiable data collected from the Innowell Fitness app. The dataset contained no personal identifiers or identifying information. In accordance with the National Health and Medical Research Council's National Statement on Ethical Conduct in Human Research (2023), Section 5.1.17(a), this research was deemed eligible for exemption from ethics review by the University of Sydney Human Research Ethics Committee as it constitutes low risk research using pre-existing, deidentified data. All participants provided consent for their data to be used for this research activity. No participants were compensated for their involvement in this research activity.

Study Population

This study was conducted using data collected from individuals from an adult population of Australians engaged in work or university education, who were invited to join the Innowell Fitness app through their respective institutions. Individuals within this study used the app between September 2021 to September 2023 (N=5933). Using the Innowell Fitness app, individuals completed a self-report questionnaire that measured their social support, personal functioning, psychological distress, sleep, physical activity, nutrition, and alcohol and other substance use. All items within the questionnaire must be answered to be recorded within the app, and thus, all questionnaires used in this analysis are complete. We used their first completed questionnaire on the app as their baseline observation. As a follow-up observation, we used their first completed questionnaire within a timeframe of 1 week to 6 months postbaseline. Individuals who reported negative minutes spent doing any physical activity were excluded (55/5933, 0.9%). Otherwise, we included all individuals who had at least 1 follow-up from 1 week to 6 months after baseline (619/5878, 11%).

Innowell Fitness App

The Innowell Fitness app is a digital mental health tool used by adults for the assessment, management, and monitoring of their mental health and well-being [22,23]. It is available on mobile and computer devices, and includes (1) self-report assessments about each domain of a person's mental fitness, (2) actionable insights about each domain of mental fitness, (3) personalized (knowledge-based) recommendations with evidence-based strategies and resources to understand and manage mental health and well-being, and (4) a goal setting and tracking tool which provides people with habit-forming activities designed to improve their mental fitness. The development of this tool involved a team of psychologists, psychiatrists, mental health research experts, and those with lived experience, who selected items that measure various components of mental health and well-being and collated relevant evidence-based strategies and resources.

The measures and domains assessed in the app include (1) social support, using 3 items from the Schuster Social Support Scale [24]; (2) personal functioning, using 3 items about educational and employment engagement and achievement [25]; (3) psychological distress, using the K6 (Kessler-6) scale for psychological distress [26]; (4) sleep, using 4 sleep items, including feeling refreshed after sleep, trouble falling asleep, and subjective energy [27]; (5) physical activity, 4 items from the International Physical Activity Questionnaire measuring time spent walking, doing moderate exercise, doing vigorous exercise, and being sedentary [28]; (6) alcohol and other substance use, using 3 items about tobacco, alcohol, and other substance use [29,30]; and (7) nutrition, using 2 items about typical composition and portion size of their diet [30]. These individual items are combined to construct domains of interest that are categorized as either "poor," "fair," or "healthy." Item and categorization details can be found in Note S1 in [Multimedia Appendix 1](#).

Statistical Analysis

Statistical modeling and analyses were performed in R (version 4.3.3; R foundation). Causal inference was performed in the structural causal modeling (SCM) framework. An SCM is described by a set of variables, a set of functions relating the variables, and a causal structure represented by a directed acyclic graph (DAG) that indicates the directionality of causal influence using arrows between random variables. In many causal inference applications, researchers encode causal assumptions within a single SCM, but we suggest that this is not appropriate in the mental health and well-being context, where the causal processes are not well known, and may vary in different contexts or for different individuals. Thus, we will marginalize (ie, Bayesian average) over the distribution of SCMs.

Marginalizing over the SCMs requires the estimation of its posterior distribution. We aimed to sample from the posterior distribution assuming a uniform prior over DAGs, only excluding DAGs with arrows that go backwards in time. Posterior sampling was achieved using an implementation of the Partition Markov Chain Monte Carlo scheme [31,32]. Partition Markov Chain Monte Carlo samples from the space of partitions, where a partition is a set of weakly ordered nodes that represents multiple DAGs. For example, consider a set of variables $\{A, B, C, D\}$ that have the dependency relations described by the DAG $A \leftarrow B \rightarrow C \rightarrow D$, then this DAG would be consistent with the partition $\{\{B\}, \{A, C\}, \{D\}\}$. The sampling procedure was run across 8 chains and checked for convergence and resolution (Note S2 in [Multimedia Appendix 1](#)). We used the Bayesian Gaussian equivalent score to retain the ordinal information of the random variables.

Simulating outcomes given a DAG is performed by constructing a Bayesian network (BN). A BN assumes nodes take categorical values and relates a variable to its parent variables using conditional probability tables. A BN was constructed per posterior sample by passing the DAG and observed data to the gRain library [33], which estimates the conditional probability tables using a maximum a posteriori estimate. Simulating an outcome given an observed baseline

state is performed by setting each baseline node with the values of the observed baseline state and then simulating the follow-up state. This corresponds to doing “nothing” below, which is shorthand for simulating a follow-up state given no intervention. The interventional do-operation is used to simulate outcomes given idealized interventions. This is performed by “mutilating” the BN by removing all edges into an interventional node, setting the interventional node state to “healthy,” and then setting the states of the baseline nodes equal to the given baseline state. Note that this intervention acts on a domain at follow-up.

We assign numerical values to outcomes, which are referred to as utilities in decision theory, to make comparisons between idealized interventions [15,34]. For our primary assumption, we assign the subutility value for outcomes as $u(\text{“poor”}=0, \text{“fair”}=0.75, \text{“healthy”}=1)$, which corresponds to a moderately risk-averse subutility function. The utility function is then an equal-weighted sum of the subutility values across domains, thus assuming no preference between domains, and ensuring that our focus is on overall well-being. The utility ranges from 0 when all domains are “poor” to 7 when all domains are “healthy.” We use the EU principle to order intervention targets, where we assume that an intervention target A is preferred to B when the EU for performing an idealized intervention on A is greater than on B . In our sensitivity analysis, we investigated a risk-neutral subutility function with $u(\text{“poor”}=0, \text{“fair”}=0.5, \text{“healthy”}=1)$, a risk-averse subutility function $u(\text{“poor”}=0, \text{“fair”}=1, \text{“healthy”}=1)$, and up-weighting psychological distress and personal functioning compared to all other domains. Further detail about the BDT framework is provided in Note S3 in Multimedia Appendix 1.

We report the average treatment effect (ATE) conditional on the baseline state. This is calculated as the difference in the EU between doing an idealized intervention compared to doing nothing.

Results

Sample Characteristics

The sample comprised 619 individuals who used the Innowell Fitness app from the original 5933 cohort. Individuals in our sample tended to have a slightly higher propensity of being in the healthy category across personal functioning (sample $n=209$, 34%; cohort $n=1663$, 28%), psychological distress (sample $n=326$, 53%; cohort $n=2818$, 48%), nutrition (sample $n=268$, 43%; cohort $n=2388$, 40%), physical activity (sample $n=466$, 75%; cohort $n=4284$, 72%), sleep (sample $n=173$, 28%; cohort $n=1611$, 27%), social support (sample $n=180$, 29%; cohort $n=1714$, 29%), and substance use (sample $n=383$, 62%; cohort $n=3471$, 59%). The median follow-up time was 55 (IQR 35-96) days, with further breakdown in Note S4 in Multimedia Appendix 1. The sample improved across a range of outcomes from baseline to follow-up where the number of individuals moving from “fair” or “poor” to “healthy” was +45 (7.3%) for sleep, +17 (2.7%) for physical activity, +28 (4.5%) for social support, –1 (0.2%) for personal functioning, +50 (8.1%) for psychological distress, 0 (0%) for substance use, and +24 (3.9%) for nutrition. Further details are provided in Table 1.

Table 1. Sample characteristics. Comparison of the analyzed sample ($n=619$; 10%) to the cohort of working individuals who have used the Innowell Fitness app ($N=5933$).

	Healthy, n (%)	Fair, n (%)	Poor, n (%)
Sleep			
Baseline (cohort)	1611 (27)	3282 (55)	1040 (18)
Baseline	173 (28)	343 (55)	103 (17)
Follow-up	218 (35)	307 (50)	94 (15)
Difference between time points	45 (7.3)	–36 (5.8)	–9 (1.5)
Physical activity			
Baseline (cohort)	4284 (72)	294 (5)	1355 (23)
Baseline	466 (75)	22 (3.5)	131 (21)
Follow-up	483 (78)	24 (3.9)	112 (18)
Difference between time points	17 (2.7)	2 (0.3)	–19 (3.1)
Social support			
Baseline (cohort)	1714 (29)	1675 (28)	2544 (43)
Baseline	180 (29)	169 (27)	267 (43)
Follow-up	208 (34)	205 (33)	203 (32)
Difference between time points	28 (4.5)	36 (5.8)	–64 (10)
Personal functioning			
Baseline (cohort)	1663 (28)	1907 (32)	2363 (40)
Baseline	209 (34)	186 (30)	224 (36)

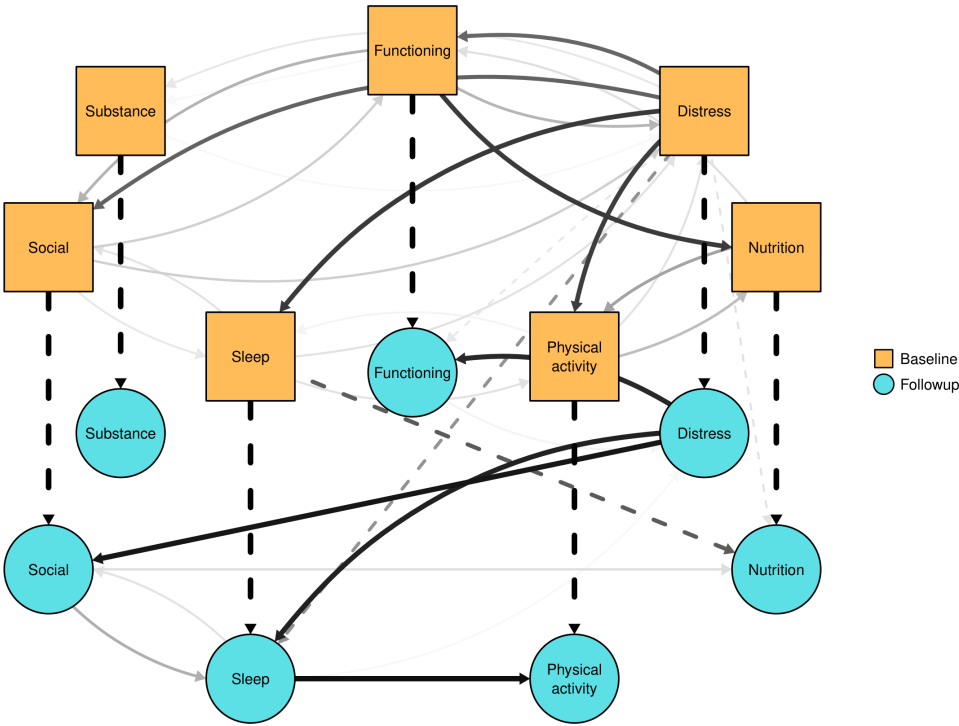
	Healthy, n (%)	Fair, n (%)	Poor, n (%)
Follow-up	208 (34)	181 (29)	230 (37)
Difference between time points	−1 (0.2)	−5 (0.8)	6 (1)
Psychological distress			
Baseline (cohort)	2818 (48)	1784 (30)	1331 (22)
Baseline	326 (53)	171 (28)	122 (20)
Follow-up	376 (61)	119 (19)	124 (20)
Difference between time points	50 (8.1)	−52 (8.2)	2 (0.3)
Substance use			
Baseline (cohort)	3471 (59)	1629 (28)	833 (14)
Baseline	383 (62)	158 (26)	78 (13)
Follow-up	383 (62)	158 (26)	78 (13)
Difference between time points	0 (0)	0 (0)	0 (0)
Nutrition			
Baseline (cohort)	2388 (40)	2106 (36)	1439 (24)
Baseline	268 (43)	213 (34)	138 (22)
Follow-up	292 (47)	206 (33)	121(20)
Difference between time points	24 (3.9)	−7 (1.1)	−17 (2.7)

Structure Learning

We now summarize the posterior distribution of DAGs (Figure 2). The contemporaneous network within baseline (ie, $A_{baseline} \rightarrow A_{followup}$) shows a dense network consistent with analysis in a partly overlapping sample [22]. There is significant uncertainty about directionality, although psychological distress was more likely to be the parent

of another domain than in the opposite direction. Specifically, psychological distress being the parent of (1) personal functioning was 64% compared to 36% in the opposite direction, (2) physical activity was 78% compared to 20%, (3) sleep was 80% compared to 20%, and (4) 65% to social support compared to 27%. Additionally, personal functioning was more likely to affect nutrition (79% compared to 21%).

Figure 2. Consensus graph summarizing the posterior distribution of DAGs. Edges between nodes with probability (p ; $p_{parent}>10\%$), with darker colors corresponding to higher probability. Dashed lines show lagged edges whereas solid lines show within time point edges. DAG: directed acyclic graph.



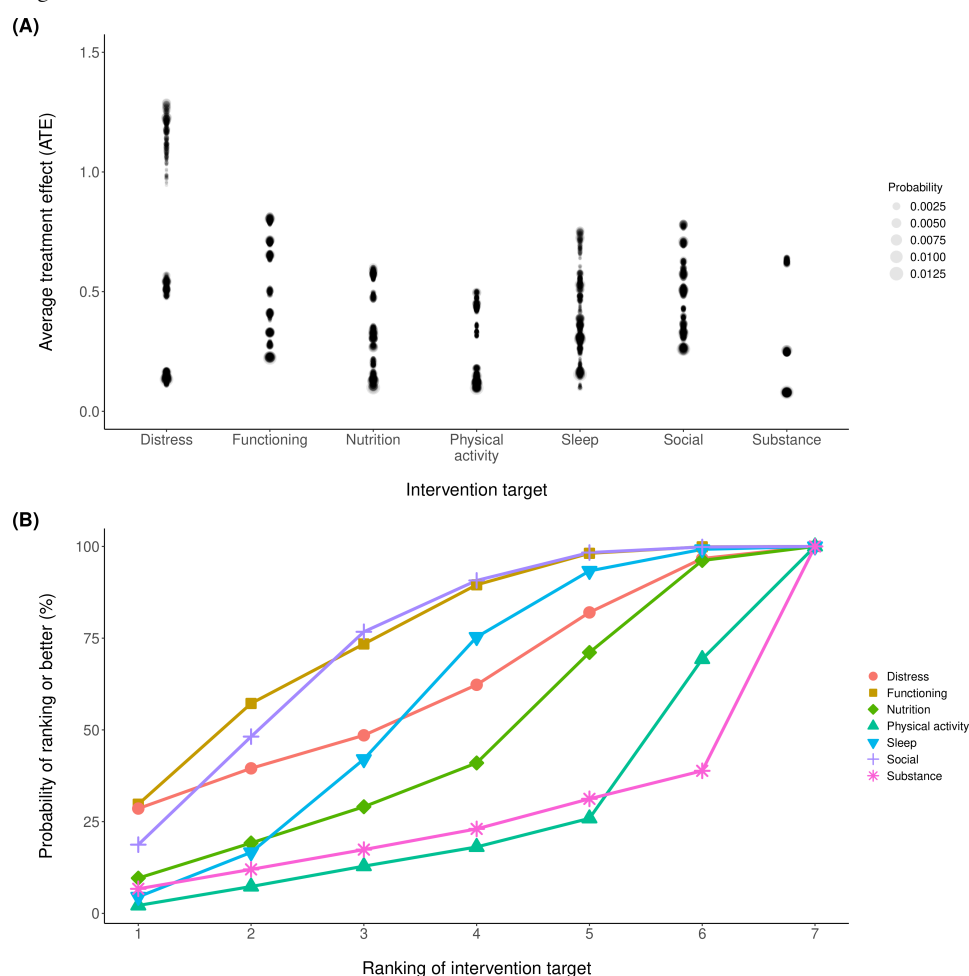
Adding the time component helps the algorithm differentiate directionality within the follow-up time point. We found the autoregressive edges (ie, $A_{baseline} \rightarrow A_{followup}$) had probability (p), $p_{parent} > 99\%$ for all domains. The lagged cross-domain edges (ie, $A_{baseline} \rightarrow B_{followup}$) are sleep to nutrition ($p_{parent}=68\%$) and potentially psychological distress to sleep ($p_{parent}=48\%$). Within the follow-up time point, we found psychological distress to personal functioning ($p_{parent}=86\%$), social support ($p_{parent}=92\%$), and sleep ($p_{parent}=88\%$), along with sleep to physical activity ($p_{parent}=95\%$). Including indirect paths, there were paths from psychological distress to personal functioning ($p_{path}=86\%$),

social support ($p_{path}=92\%$), sleep ($p_{path}=88\%$), and physical activity ($p_{path}=86\%$). Further details are in Note S5 in [Multimedia Appendix 1](#).

Treatment Effects

Intervening on psychological distress is capable of the greatest ATE, with ATE > 1 (equivalent to transitioning from “poor” to “healthy” for 1 domain), when psychological distress itself was adjusted from “poor” to “healthy” while also affecting other domains. Interventions on other domains resulted in ATE less than 1 due to effects being primarily isolated to that domain ([Figure 3A](#)).

Figure 3. Treatment effects and recommendation targets. Panel (A) shows average treatment effects with point size increasing with an empirical probability of the baseline state. These are converted to recommendations represented in panel (B), where the probability for the rank of each intervention target is marginalized over the baseline states.



Decision Analysis

We estimated a preference ranking for each domain ([Figure 3B](#)). The proportion that a domain is the optimal intervention weighted per the probability of the baseline states is personal functioning ($p_{opt}=30\%$), psychological distress ($p_{opt}=29\%$), social support ($p_{opt}=18\%$), nutrition ($p_{opt}=9.6\%$), substance use ($p_{opt}=6.7\%$), sleep ($p_{opt}=4.5\%$), and physical activity ($p_{opt}=2.2\%$).

We also report the proportion that a domain would be recommended in a system that presented the top N interventional targets. For example, for a system that reported the

top three intervention targets the proportion that each domain would be displayed would be social support (probability of being recommended, $p_{rec}=77\%$), personal functioning ($p_{rec}=73\%$), psychological distress ($p_{rec}=49\%$), sleep ($p_{rec}=42\%$), nutrition ($p_{rec}=29\%$), substance use ($p_{rec}=17\%$), and physical activity ($p_{rec}=13\%$).

Optimal interventional targets can be further illuminated by examples. The most common baseline presentation is everything is “healthy,” where the optimal interventional target was social support (EU 6.345, SE 0.034), followed by personal functioning (EU 6.310, SE 0.035), which were

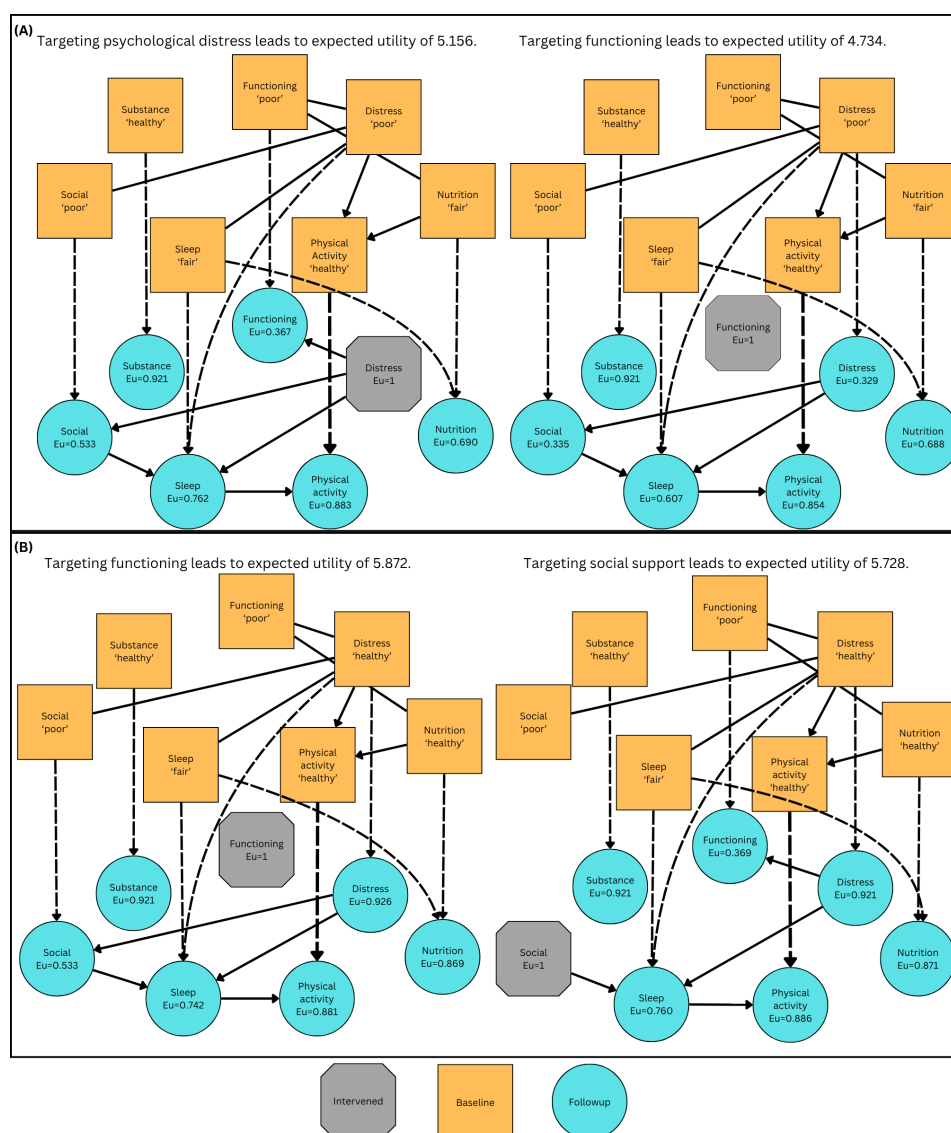
greater than doing nothing (EU 6.087, SE 0.033). This is due to a regression to the mean effect, where social support (expected subutility [Eu] 0.764, SE 0.010) and personal functioning (Eu 0.784, SE 0.010) tend to revert to unhealthy states with greater probability than other domains, which all had $Eu > 0.9$.

Domains that are less healthy than all other domains are typically the optimal interventional target. However, when multiple domains including psychological distress are “fair” with either nutrition or physical activity as “poor,” psychological distress was the optimal target. For example, psychological distress is the optimal intervention target for the state (functioning=“fair,” psychological

distress=“fair,” nutrition=“fair,” physical activity=“poor,” sleep=“fair,” social=“fair,” substance=“healthy”), as it affects multiple domains.

Psychological distress was the optimal intervention if it was equally unhealthy to any other domain. For example, assuming the baseline state (functioning=“poor,” psychological distress=“poor,” nutrition=“fair,” physical activity=“healthy,” sleep=“fair,” social=“poor,” substance=“healthy”), the optimal intervention target is psychological distress (EU 5.156, SE 0.115), rather than personal functioning (EU 4.734, SE 0.087), or social support (EU 4.638, SE 0.122) despite those domains also being poor (Figure 4A).

Figure 4. Comparison of predicted outcomes and utilities following interventions for different baseline presentations shown in panel (A) and (B). We show edges in the maximum a posteriori completed partially DAG, where undirected edges correspond to cases where DAGs with edges in either direction have the same posterior probability. Dashed lines show lagged edges whereas solid lines show within time point edges. DAG: directed acyclic graph; Eu: expected subutility.



When psychological distress was “healthy” and multiple domains were unhealthy, the interventional target was less certain. For example, for the state (functioning=“poor,” psychological distress=“healthy,” nutrition=“healthy,”

physical activity=“healthy,” sleep=“fair,” social=“poor,” substance=“healthy”), the optimal interventional target was personal functioning (EU 5.872, SE 0.059) instead of social support (EU 5.728, SE 0.067). This is for similar reasons that

the unhealthy personal functioning state persists with greater probability than social support (Figure 4B).

We tested the sensitivity of preference rankings to utility function assumptions. These different assumptions could represent different expert or individual-level outcome preferences. Assuming a risk-neutral subutility function with $u(\text{"poor"}=0, \text{"fair"}=0.5, \text{"healthy"}=1)$, the optimal intervention targets were psychological distress ($p_{\text{opt}}=28\%$), personal functioning ($p_{\text{opt}}=27\%$), social support ($p_{\text{opt}}=18\%$), sleep ($p_{\text{opt}}=11\%$), nutrition ($p_{\text{opt}}=7.4\%$), substance use ($p_{\text{opt}}=7.1\%$), and physical activity ($p_{\text{opt}}=1.4\%$). We then tested for a highly risk-averse subutility function $u(\text{"poor"}=0, \text{"fair"}=1, \text{"healthy"}=1)$, where we found psychological distress ($p_{\text{opt}}=28\%$), personal functioning ($p_{\text{opt}}=28\%$), social support ($p_{\text{opt}}=24\%$), nutrition ($p_{\text{opt}}=8.2\%$), substance use ($p_{\text{opt}}=6.7\%$), physical activity ($p_{\text{opt}}=4.3\%$), and sleep ($p_{\text{opt}}=0.9\%$). We also increased the weighting of psychological distress and personal functioning using weakly ordered domain-ranking preferences, where we found distress ($p_{\text{opt}}=39\%$), personal functioning ($p_{\text{opt}}=37\%$), social support ($p_{\text{opt}}=11\%$), nutrition ($p_{\text{opt}}=5.7\%$), substance use ($p_{\text{opt}}=4.3\%$), sleep ($p_{\text{opt}}=1.9\%$), and physical activity ($p_{\text{opt}}=1.2\%$). Further details are in Note S6 in [Multimedia Appendix 1](#).

Discussion

Principal Findings

We developed a CAIRS that can be incorporated into digital mental health care tools. Our primary contribution is to show how causal effects and their uncertainty, estimated using causal AI, can be incorporated into automated RSs. We also explored how those recommendations would change per expert or individual-level outcome preferences.

Using our example, we found the optimal interventional target is a function of an individual's presentation (which in our example is their baseline state), the noninterventional transition from baseline to follow-up, the causal effects of the intervention on itself and other domains, and the utility function. To summarize our results, the optimal interventional target assuming the default utility function was (1) the unhealthiest baseline domain with some exceptions where psychological distress is more effective than intervening on "poor" nutrition or physical activity, (2) psychological distress when it is equally or more unhealthy than other domains, or (3) the domain that is more likely to transition to or persist in an unhealthy state.

Personalizing Recommendations

CAIRS provides mechanisms to personalize recommendations. In our example, conditioning recommendations on an individual's baseline presentation leads to recommending interventions consistent with an individual's current or predicted future problem domains. Intervening on the unhealthiest domain would be the conclusion of most systems. Similarly, domains that are more likely to persist in or transition to unhealthy states suggest that intervention

is required. While many RSs may implement this latter step if known, we note that this finding was not something we considered before performing the analysis and is not incorporated into the current Innowell Fitness knowledge-based recommendations, suggesting added value from our analysis.

We also explored personalization of recommendations to an individual's outcome preferences. Comparing multidimensional outcomes per expert or individual preferences is a challenging task that BDT provides a framework to computationally encode within an RS. Some possibilities of varying outcome preferences were explored by adjusting the subutility function to account for different risk-aversity and changing the domain weightings, which corresponded to slight recommendation changes. Further adjustments in the utility function are possible, and this utility framework will become increasingly important as we increase the number of domains or investigate symptoms, as it is unlikely that all relevant domains or symptoms would be considered equally important by all individuals or experts.

Incorporating personalized outcome preferences in a deployed app would require a utility elicitation mechanism. There are many utility elicitation methods available, some of which are not overly burdensome, such as eliciting weakly ordered preference for domains [35], while BDT provides mechanisms to stop utility elicitation when there is enough information to make a recommendation [36]. With that said, our analysis does not address the feasibility of eliciting individualized utilities in real-world applications, which should be considered for further research. Further preferences (eg, associated costs or interventional preferences) would also improve the personalization of the recommendations.

BDT could find applicability elsewhere within digital health applications. It could also be applied to other causal inference or predictive frameworks, such as undirected or dynamic networks, where intervention targets have been explored [37-42], but often assume that all domains or symptoms are equally important and typically do not marginalize out uncertainties. We may also want to balance mental health and well-being outcomes with other considerations, such as user engagement, which is often poor in digital mental health tools [43-45]. We keep this out of the current scope, only noting that utility functions are a broad framework that can be used to incorporate a range of aspects for optimization.

Recommending Interventions

Within the clinical setting, CAIRS will aid the implementation of highly personalized care [4]. In these frameworks, the effects between multiple dimensions are considered to inform the processes that affect an individual's mental ill-health and improve interventional decision-making. These approaches are often supplemented by ongoing measurement-based care to monitor an individual's progress [46], with several digital mental health tools that support implementation [23]. While these tools have aided highly personalized care across domains, they rarely provide multidimensional interventional insights beyond simple knowledge-based approaches. CAIRS

could enhance these tools by providing a data-driven understanding of the causal processes between domains, along with simplifying decisions by filtering interventions or intervention targets across several domains.

Furthermore, in cases where interventions are well recorded, it would be possible to include them in the SCMs, such that they could be recommended algorithmically. Although, causal discovery with interventions can be difficult as they often influence multiple variables [47] and the ethical challenges of recommending interventions will need to be considered. Including clinical symptoms and other relevant clinical variables would also improve the specification of interventions. Additionally, the SCMs will need to be sampled within each new setting.

An advantage of our approach is the improved explainability of the recommendations compared to most AI algorithms. Causal insight can be communicated to clinicians or individuals by displaying the SCMs graphically, similar to what is shown in Figure 4. Training would be required for clinicians to fully understand the outputs of CAIRS. However, clinicians undergo causal reasoning during case formulation similar to that in CAIRS [48], and thus, we suspect it will be feasible for clinicians to understand its recommendations and insights.

In our example, targeting psychological distress should be preferred over other equally unhealthy domains in this general population cohort of adults engaged in the workforce or university education. This result is due to the causal effects that psychological distress has on personal functioning, social support, sleep, and physical activity (mediated by sleep). This is consistent with current understanding that mental ill-health affects multiple domains, including work concerning absenteeism and productivity [49,50], sleeping patterns, for example, due to rumination [51,52], and social connection due to social anxiety [53].

This has interventional implications for individuals within this cohort. Psychological distress as measured using the K6 scale has been shown to have very good accuracy at identifying serious mental ill-health in general populations, typically defined as any mental disorder meeting diagnostic threshold within the following 12 months [54,55]. Therefore, CAIRS could identify and direct individuals from this general population to targeted clinical interventions, such as cognitive behavioral therapies, which have been shown to be efficacious for a range of serious mental illnesses within many contexts [56], including internet-delivered therapies [57,58]. The advantage of CAIRS rather than solely using a K6 score is to balance psychological distress against other considerations, including other problem domains and any other individualized preferences or costs. Thus, CAIRS could reduce clinical resources in the assignment of individuals to relevant therapies, which could lead to improved access to such therapies.

Comparisons to Prior Work

RSs within digital mental health and well-being contexts have been used to recommend content to individuals [13,14].

These algorithms recommend content to individuals using the resemblance of user input, expert-level knowledge-based systems, or predictive AI. While these algorithms may be useful in some contexts, they do not appropriately consider outcomes under interventions, which we argue is fundamental to recommending interventions, and is incorporated in CAIRS.

Components of our approach have been explored previously. Causal AI with structure learning has been used for epidemiological purposes in mental health [59-61]. These analyses have found a wide array of dependencies between variables, both cross-sectionally and longitudinally. Those analyses have been used to estimate average causal effects using a counterfactual query of the SCM, which provides a broad understanding of interventional targets. Our approach improves on this by leveraging personalized information, including their current state and outcome preferences, which is required to make individual-level decisions.

BDT principles have also been applied for health care decision problems [62-64]. These applications typically assume a causal structure (often implicitly rather than explicitly through a DAG), which may not be appropriate for applications in mental health and well-being, where the causal structure may vary. Therefore, combining causal structure learning and BDT to account for the uncertainty in the SCM provides a novel approach for applications in digital mental health.

Analytical Limitations

First, our results rely on causal interpretations of the inferred DAGs. Assumptions must hold for this to be true, including that we have all relevant confounders and colliders [65]. Missing variables can lead to incorrectly inferred SCMs, which would lead to incorrect causal effects and recommendations. For the causal effect estimations to be valid other factors such as sociodemographics, historical values before baseline, or other domains have a negligible causal effect on follow-up variables beyond the effect that they have on the observed variables. In our example, a wide array of domains that could feasibly interact were investigated, and thus, this assumption may be reasonable. However, these assumptions may not hold and should be tested more thoroughly, including improving the recording of potential confounders within the Innwell Fitness app.

Second, transferability of our results from our example to other cohorts is also limited. The relevant variables should be tailored to that cohort and the SCMs will have to be learned within that context. With that said, we suspect that the CAIRS approach will be broadly applicable to a wide array of contexts.

Third, we acknowledge that a data-driven approach to causal discovery, as we have attempted, may not be adequate to accurately estimate SCMs. Although we suspect that the full SCM is unlikely to be well known in the mental health and well-being context, some aspects of the causal structure may be known, or at least highly probable, given knowledge about those constructs within a specific context or for an

individual. Bayesian inference for causal structures allows for this prior knowledge to be incorporated into the structure learning procedure [16]. Therefore, future iterations of CAIRS should incorporate prior causal knowledge into the procedure.

Fourth, the true causal paths are probably more complex than suggested in this work, as more causal effects have been found in other contexts between the domains that we have studied. The data is likely underpowered to determine all causal paths. However, the BDT framework used within this analysis accounts for uncertain variables, including the SCM, which should allow for appropriate recommendations given the information available in the data.

Lastly, within time point, cyclic paths will be missing due to the acyclic constraint on the causal structure. The acyclic limitation reduces the cross-domain causal effects within a given time point, which may underestimate the effects of interventions in some cases. Accounting for this appropriately will require the incorporation of recent methodological

developments [42], which allow for alternative representations of SCMs to account for cycles.

Conclusions

Interventional recommendation algorithms require careful incorporation of causal considerations and decision-making principles. Important aspects are the estimation of outcomes under interventions requiring causal modeling, the assignment of appropriate utilities to align recommendations with outcome preferences, and consideration of uncertainty. These considerations can be incorporated into computational models using causal AI and BDT. In our example application aimed at a general population, we show that CAIRS can recommend reasonable intervention targets across multiple domains, which we argue could be used to assign intervention targets, such as directing individuals to clinical therapies. Therefore, this approach has the potential to streamline recommendations for complex decisions that are common in digital mental health care.

Acknowledgments

This work was supported by the Medical Research Future Fund National Critical Research Infrastructure Grant (MRFCRI000279) and National Health and Medical Research Council (NHMRC) Australia Fellowship (no. 511921 awarded to IBH). MV was supported by philanthropic funding from the Johnston Fellowship and from other donors who are families affected by mental illness who wish to remain anonymous. IBH is supported by an NHMRC L3 Investigator Grant (GNT2016346). JJC was supported by an NHMRC Emerging Leadership Fellowship (GNT2008196). FI was supported by an NHMRC EL1 Investigator Grant (GNT2018157).

Data Availability

Data used for this study is available from the corresponding author on reasonable request.

Authors' Contributions

MV and FI conceptualized this study. MV, VA, SC, and RM developed the software and methodologies. MV, VA, and FI wrote the first draft. Draft reviewing and editing were performed by IBH, SC, RM, JS, JJC, BO, AP, and SM. All authors had full access to the data in this study and had final responsibility for the decision to submit for publication. MV and VA have directly accessed and verified the data.

Conflicts of Interest

IBH is the codirector of Health and Policy at the Brain and Mind Centre (BMC), University of Sydney, Australia. The BMC operates an early intervention youth service at Camperdown under contract to headspace. IBH has previously led community-based and pharmaceutical industry-supported (Wyeth, Eli Lilly, Servier, Pfizer, AstraZeneca, and Janssen Cilag) projects focused on the identification and better management of anxiety and depression. IBH is the chief scientific adviser to, and a 3.2% equity shareholder in, Innowell Pty Ltd, which aims to transform mental health services through the use of innovative technologies. All other authors declare no conflicts of interest. All other authors declare no financial or nonfinancial competing interests.

Multimedia Appendix 1

Additional information and results.

[\[DOCX File \(Microsoft Word File\), 999 KB-Multimedia Appendix 1\]](#)

References

1. Scheffer M, Bockting CL, Borsboom D, et al. A dynamical systems view of psychiatric disorders-practical implications: a review. *JAMA Psychiatry*. Jun 1, 2024;81(6):624-630. [doi: [10.1001/jamapsychiatry.2024.0228](https://doi.org/10.1001/jamapsychiatry.2024.0228)] [Medline: [38568618](https://pubmed.ncbi.nlm.nih.gov/38568618/)]
2. Bringmann L, Helmich M, Eronen M, Voelkle M. *Complex Systems Approaches to Psychopathology*. Oxford Academic; 2023. [doi: [10.1093/med-psych/9780197542521.003.0005](https://doi.org/10.1093/med-psych/9780197542521.003.0005)]
3. Borsboom D. A network theory of mental disorders. *World Psychiatry*. Feb 2017;16(1):5-13. [doi: [10.1002/wps.20375](https://doi.org/10.1002/wps.20375)] [Medline: [28127906](https://pubmed.ncbi.nlm.nih.gov/28127906/)]

4. Hickie IB, Scott EM, Cross SP, et al. Right care, first time: a highly personalised and measurement-based care model to manage youth mental health. *Med J Aust*. Nov 2019;211 Suppl 9(S9):S3-S46. [doi: [10.5694/mja2.50383](https://doi.org/10.5694/mja2.50383)] [Medline: [31679171](https://pubmed.ncbi.nlm.nih.gov/31679171/)]
5. McGorry PD, Mei C, Dalal N, et al. The Lancet Psychiatry Commission on youth mental health. *Lancet Psychiatry*. Sep 2024;11(9):731-774. [doi: [10.1016/S2215-0366\(24\)00163-9](https://doi.org/10.1016/S2215-0366(24)00163-9)] [Medline: [39147461](https://pubmed.ncbi.nlm.nih.gov/39147461/)]
6. Macintyre A, Ferris D, Gonçalves B, Quinn N. What has economics got to do with it? The impact of socioeconomic factors on mental health and the case for collective action. *Palgrave Commun*. 2018;4(1):10. [doi: [10.1057/s41599-018-0063-2](https://doi.org/10.1057/s41599-018-0063-2)]
7. Parmigiani G, Inoue L. *Decision Theory: Principles and Approaches*. John Wiley & Sons, Ltd; 2009. [doi: [10.5860/choice.47-4475](https://doi.org/10.5860/choice.47-4475)]
8. Gomez-Urbe CA, Hunt N. The Netflix recommender system: algorithms, business value, and innovation. *ACM Trans Manag Inf Syst*. 2015;6(4):13. [doi: [10.1145/2843948](https://doi.org/10.1145/2843948)]
9. De Croon R, Van Houdt L, Htun NN, Štiglic G, Vanden Abeele V, Verbert K. Health recommender systems: systematic review. *J Med Internet Res*. Jun 29, 2021;23(6):e18035. [doi: [10.2196/18035](https://doi.org/10.2196/18035)] [Medline: [34185014](https://pubmed.ncbi.nlm.nih.gov/34185014/)]
10. Etemadi M, Abkenar SB, Ahmadzadeh A, et al. A systematic review of healthcare recommender systems: open issues, challenges, and techniques. *Expert Syst Appl*. Mar 2023;213:118823. [doi: [10.1016/j.eswa.2022.118823](https://doi.org/10.1016/j.eswa.2022.118823)]
11. Chaturvedi A, Aylward B, Shah S, et al. Content recommendation systems in web-based mental health care: real-world application and formative evaluation. *JMIR Form Res*. Jan 19, 2023;7:e38831. [doi: [10.2196/38831](https://doi.org/10.2196/38831)] [Medline: [36656628](https://pubmed.ncbi.nlm.nih.gov/36656628/)]
12. Valentine L, D'Alfonso S, Lederman R. Recommender systems for mental health apps: advantages and ethical challenges. *AI Soc*. Jan 17, 2022;38(4):1-12. [doi: [10.1007/s00146-021-01322-w](https://doi.org/10.1007/s00146-021-01322-w)] [Medline: [35068708](https://pubmed.ncbi.nlm.nih.gov/35068708/)]
13. Siriaraya P, Suzuki K, Nakajima S. Utilizing collaborative filtering to recommend opportunities for positive affect in daily life. The 4th Workshop on Health Recommender Systems co-located with ACM RecSys 2019; 2019. URL: <https://healthrecsys.github.io/2019/HealthRec2019Panote.pdf>
14. Rohani DA, Lopategui AQ, Tuxen N, Faurholt-Jepsen M, Kessing LV, Bardram JE. MUBS: a personalized recommender system for behavioral activation in mental health. Presented at: CHI '20; Apr 25-30, 2020; Honolulu, HI, United States. URL: <https://dl.acm.org/doi/proceedings/10.1145/3313831> [Accessed 2025-06-07] [doi: [10.1145/3313831.3376879](https://doi.org/10.1145/3313831.3376879)]
15. Pearl J. *Causality: Models, Reasoning, and Inference*. Cambridge University Press; 2000. [doi: [10.1093/bjps/52.3.613](https://doi.org/10.1093/bjps/52.3.613)]
16. Cripps S. Artificial and human intelligence. *J Proc R Society N S W*. 2024;157(1):109-118. [doi: [10.3316/informit.T2024062800017390098901872](https://doi.org/10.3316/informit.T2024062800017390098901872)]
17. Aboujaoude E, Gega L, Parish MB, Hilty DM. Editorial: digital interventions in mental health: current status and future directions. *Front Psychiatry*. 2020;11(February):111. [doi: [10.3389/fpsyt.2020.00111](https://doi.org/10.3389/fpsyt.2020.00111)] [Medline: [32174858](https://pubmed.ncbi.nlm.nih.gov/32174858/)]
18. Eisenstadt M, Liverpool S, Infanti E, Ciuvat RM, Carlsson C. Mobile apps that promote emotion regulation, positive mental health, and well-being in the general population: systematic review and meta-analysis. *JMIR Ment Health*. Nov 8, 2021;8(11):e31170. [doi: [10.2196/31170](https://doi.org/10.2196/31170)] [Medline: [34747713](https://pubmed.ncbi.nlm.nih.gov/34747713/)]
19. Woodward K, Kanjo E, Brown DJ, et al. Beyond mobile apps: a survey of technologies for mental well-being. *IEEE Trans Affective Comput*. 2022;13(3):1216-1235. [doi: [10.1109/TAFFC.2020.3015018](https://doi.org/10.1109/TAFFC.2020.3015018)]
20. Prosperi M, Guo Y, Sperrin M, et al. Causal inference and counterfactual prediction in machine learning for actionable healthcare. *Nat Mach Intell*. 2020;2(7):369-375. [doi: [10.1038/s42256-020-0197-y](https://doi.org/10.1038/s42256-020-0197-y)]
21. Sanchez P, Voisey JP, Xia T, Watson HI, O'Neil AQ, Tsaftaris SA. Causal machine learning for healthcare and precision medicine. *R Soc Open Sci*. Aug 2022;9(8):220638. [doi: [10.1098/rsos.220638](https://doi.org/10.1098/rsos.220638)] [Medline: [35950198](https://pubmed.ncbi.nlm.nih.gov/35950198/)]
22. Iorfino F, Varidel M, Capon W, et al. Quantifying the interrelationships between physical, social, and cognitive-emotional components of mental fitness using digital technology. *NPJ Ment Health Res*. Jul 8, 2024;3(1):36. [doi: [10.1038/s44184-024-00078-7](https://doi.org/10.1038/s44184-024-00078-7)] [Medline: [38977903](https://pubmed.ncbi.nlm.nih.gov/38977903/)]
23. Iorfino F, Cross SP, Davenport T, et al. A digital platform designed for youth mental health services to deliver personalized and measurement-based care. *Front Psychiatry*. 2019;10:595. [doi: [10.3389/fpsyt.2019.00595](https://doi.org/10.3389/fpsyt.2019.00595)] [Medline: [31507465](https://pubmed.ncbi.nlm.nih.gov/31507465/)]
24. Schuster TL, Kessler RC, Aseltine RH. Supportive interactions, negative interactions, and depressed mood. *Am J Community Psychol*. Jun 1990;18(3):423-438. [doi: [10.1007/BF00938116](https://doi.org/10.1007/BF00938116)] [Medline: [2264558](https://pubmed.ncbi.nlm.nih.gov/2264558/)]
25. Ustun TB, Kostanjsek N, Chatterji S, Rehm J. *WHO Short Disability Assessment Schedule (WHO DAS-s)*. Cambridge University Press; 2010. [doi: [10.1017/CBO9780511759055.008](https://doi.org/10.1017/CBO9780511759055.008)]
26. Kessler RC, Andrews G, Colpe LJ, et al. Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychol Med*. Aug 2002;32(6):959-976. [doi: [10.1017/S0033291702006074](https://doi.org/10.1017/S0033291702006074)]

27. Buysse DJ, Reynolds CF III, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh sleep quality index: a new instrument for psychiatric practice and research. *Psychiatry Res*. May 1989;28(2):193-213. [doi: [10.1016/0165-1781\(89\)90047-4](https://doi.org/10.1016/0165-1781(89)90047-4)]
28. Craig CL, Marshall AL, Sjöström M, et al. International physical activity questionnaire: 12-country reliability and validity. *Med Sci Sports Exerc*. Aug 2003;35(8):1381-1395. [doi: [10.1249/01.MSS.0000078924.61453.FB](https://doi.org/10.1249/01.MSS.0000078924.61453.FB)] [Medline: [12900694](https://pubmed.ncbi.nlm.nih.gov/12900694/)]
29. Humeniuk R, Ali R, Babor TF, et al. Validation of the Alcohol, Smoking and Substance Involvement Screening Test (ASSIST). *Addiction*. Jun 2008;103(6):1039-1047. [doi: [10.1111/j.1360-0443.2007.02114.x](https://doi.org/10.1111/j.1360-0443.2007.02114.x)] [Medline: [18373724](https://pubmed.ncbi.nlm.nih.gov/18373724/)]
30. Eat for health: Australian dietary guidelines. National Health and Medical Research Council. 2013. URL: <https://www.nhmrc.gov.au/adg> [Accessed 2025-05-30]
31. Varidel M. CIA: learn and apply directed acyclic graphs for causal inference. 2024. URL: <https://cran.r-project.org/web/packages/cia/cia.pdf> [Accessed 2025-05-30]
32. Kuipers J, Moffa G. Partition MCMC for inference on acyclic digraphs. *J Am Stat Assoc*. Jan 2, 2017;112(517):282-299. [doi: [10.1080/01621459.2015.1133426](https://doi.org/10.1080/01621459.2015.1133426)]
33. Højsgaard SH. Graphical Independence Networks with the gRain Package for R. *J Stat Softw*. ;46(10). [doi: [10.18637/jss.v046.i10](https://doi.org/10.18637/jss.v046.i10)]
34. Koller D, Friedman N. Probabilistic Graphical Models: Principles and Techniques. MIT press; 2009.
35. Alinezhad A, Khalili J. New Methods and Applications in Multiple Attribute Decision Making (MADM). Springer Cham; 2019. [doi: [10.1007/978-3-030-15009-9](https://doi.org/10.1007/978-3-030-15009-9)]
36. Chajewska U, Koller D, Parr R. Making rational decisions using adaptive utility elicitation. Presented at: Proceedings of the Seventeenth National Conference on Artificial Intelligence and Twelfth Conference on Innovative Applications of Artificial Intelligence; Jul 30 to Aug 3, 2000:363-369; Austin, TX.
37. Isvoranu AM, Abdin E, Chong SA, Vaingankar J, Borsboom D, Subramaniam M. Extended network analysis: from psychopathology to chronic illness. *BMC Psychiatry*. Feb 27, 2021;21(1):119. [doi: [10.1186/s12888-021-03128-y](https://doi.org/10.1186/s12888-021-03128-y)] [Medline: [33639891](https://pubmed.ncbi.nlm.nih.gov/33639891/)]
38. Lunansky G, Naberman J, van Borkulo CD, Chen C, Wang L, Borsboom D. Intervening on psychopathology networks: evaluating intervention targets through simulations. *Methods*. Aug 2022;204(October 2021):29-37. [doi: [10.1016/j.ymeth.2021.11.006](https://doi.org/10.1016/j.ymeth.2021.11.006)] [Medline: [34793976](https://pubmed.ncbi.nlm.nih.gov/34793976/)]
39. Roefs A, Fried EI, Kindt M, et al. A new science of mental disorders: using personalised, transdiagnostic, dynamical systems to understand, model, diagnose and treat psychopathology. *Behav Res Ther*. Jun 2022;153(March):104096. [doi: [10.1016/j.brat.2022.104096](https://doi.org/10.1016/j.brat.2022.104096)] [Medline: [35500541](https://pubmed.ncbi.nlm.nih.gov/35500541/)]
40. McNally RJ. Network analysis of psychopathology: controversies and challenges. *Annu Rev Clin Psychol*. May 7, 2021;17:31-53. [doi: [10.1146/annurev-clinpsy-081219-092850](https://doi.org/10.1146/annurev-clinpsy-081219-092850)] [Medline: [33228401](https://pubmed.ncbi.nlm.nih.gov/33228401/)]
41. Ryan O, Hamaker EL. Time to intervene: a continuous-time approach to network analysis and centrality. *Psychometrika*. Mar 2022;87(1):214-252. [doi: [10.1007/s11336-021-09767-0](https://doi.org/10.1007/s11336-021-09767-0)] [Medline: [34165691](https://pubmed.ncbi.nlm.nih.gov/34165691/)]
42. Bongers S, Forré P, Peters J, Mooij JM. Foundations of structural causal models with cycles and latent variables. *Ann Statist*. 2021;49(5):2885-2915. [doi: [10.1214/21-AOS2064](https://doi.org/10.1214/21-AOS2064)]
43. Torous J, Nicholas J, Larsen ME, Firth J, Christensen H. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. *Evid Based Ment Health*. Aug 2018;21(3):116-119. [doi: [10.1136/eb-2018-102891](https://doi.org/10.1136/eb-2018-102891)] [Medline: [29871870](https://pubmed.ncbi.nlm.nih.gov/29871870/)]
44. Baumel A, Muench F, Edan S, Kane JM. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. *J Med Internet Res*. Sep 25, 2019;21(9):e14567. [doi: [10.2196/14567](https://doi.org/10.2196/14567)] [Medline: [31573916](https://pubmed.ncbi.nlm.nih.gov/31573916/)]
45. Lipschitz JM, Pike CK, Hogan TP, Murphy SA, Burdick KE. The engagement problem: a review of engagement with digital mental health interventions and recommendations for a path forward. *Curr Treat Options Psychiatry*. Sep 2023;10(3):119-135. [doi: [10.1007/s40501-023-00297-3](https://doi.org/10.1007/s40501-023-00297-3)] [Medline: [38390026](https://pubmed.ncbi.nlm.nih.gov/38390026/)]
46. Lutz W, Schwartz B, Delgadillo J. Measurement-based and data-informed psychological therapy. *Annu Rev Clin Psychol*. May 9, 2022;18:71-98. [doi: [10.1146/annurev-clinpsy-071720-014821](https://doi.org/10.1146/annurev-clinpsy-071720-014821)] [Medline: [34910567](https://pubmed.ncbi.nlm.nih.gov/34910567/)]
47. Eronen MI. Causal discovery and the problem of psychological interventions. *New Ideas Psychol*. Dec 2020;59(June 2019):100785. [doi: [10.1016/j.newideapsych.2020.100785](https://doi.org/10.1016/j.newideapsych.2020.100785)]
48. Persons JB. The Case Formulation Approach to Cognitive-Behavior Therapy. Guilford Press; 2008.
49. Kelloway EK, Dimoff JK, Gilbert S. Mental health in the workplace. *Annu Rev Organ Psychol Organ Behav*. Jan 23, 2023;10(1):363-387. [doi: [10.1146/annurev-orgpsych-120920-050527](https://doi.org/10.1146/annurev-orgpsych-120920-050527)]
50. Ferreira AI, da Costa Ferreira P, Cooper CL, Oliveira D. How daily negative affect and emotional exhaustion correlates with work engagement and presenteeism-constrained productivity. *Int J Stress Manag*. 2019;26(3):261-271. [doi: [10.1037/str0000114](https://doi.org/10.1037/str0000114)]

51. Guastella AJ, Moulds ML. The impact of rumination on sleep quality following a stressful life event. *Pers Individ Dif*. Apr 2007;42(6):1151-1162. [doi: [10.1016/j.paid.2006.04.028](https://doi.org/10.1016/j.paid.2006.04.028)]
52. Clancy F, Prestwich A, Caperon L, Tsipa A, O'Connor DB. The association between worry and rumination with sleep in non-clinical populations: a systematic review and meta-analysis. *Health Psychol Rev*. Dec 2020;14(4):427-448. [doi: [10.1080/17437199.2019.1700819](https://doi.org/10.1080/17437199.2019.1700819)] [Medline: [31910749](https://pubmed.ncbi.nlm.nih.gov/31910749/)]
53. Teo AR, Lerrigo R, Rogers MAM. The role of social isolation in social anxiety disorder: a systematic review and meta-analysis. *J Anxiety Disord*. May 2013;27(4):353-364. [doi: [10.1016/j.janxdis.2013.03.010](https://doi.org/10.1016/j.janxdis.2013.03.010)] [Medline: [23746493](https://pubmed.ncbi.nlm.nih.gov/23746493/)]
54. Kessler RC, Barker PR, Colpe LJ, et al. Screening for serious mental illness in the general population. *Arch Gen Psychiatry*. Feb 1, 2003;60(2):184. [doi: [10.1001/archpsyc.60.2.184](https://doi.org/10.1001/archpsyc.60.2.184)]
55. Kessler RC, Green JG, Gruber MJ, et al. Screening for serious mental illness in the general population with the K6 screening scale: results from the WHO World Mental Health (WMH) survey initiative. *Int J Methods Psychiatr Res*. Jun 2010;19 Suppl 1(Suppl 1):4-22. [doi: [10.1002/mpr.310](https://doi.org/10.1002/mpr.310)] [Medline: [20527002](https://pubmed.ncbi.nlm.nih.gov/20527002/)]
56. Fordham B, Sugavanam T, Edwards K, et al. The evidence for cognitive behavioural therapy in any condition, population or context: a meta-review of systematic reviews and panoramic meta-analysis. *Psychol Med*. Jan 2021;51(1):21-29. [doi: [10.1017/S0033291720005292](https://doi.org/10.1017/S0033291720005292)] [Medline: [33455594](https://pubmed.ncbi.nlm.nih.gov/33455594/)]
57. Etzelmueller A, Vis C, Karyotaki E, et al. Effects of internet-based cognitive behavioral therapy in routine care for adults in treatment for depression and anxiety: systematic review and meta-analysis. *J Med Internet Res*. Aug 31, 2020;22(8):e18100. [doi: [10.2196/18100](https://doi.org/10.2196/18100)] [Medline: [32865497](https://pubmed.ncbi.nlm.nih.gov/32865497/)]
58. Karyotaki E, Efthimiou O, Miguel C, et al. Internet-based cognitive behavioral therapy for depression: a systematic review and individual patient data network meta-analysis. *JAMA Psychiatry*. Apr 1, 2021;78(4):361-371. [doi: [10.1001/jamapsychiatry.2020.4364](https://doi.org/10.1001/jamapsychiatry.2020.4364)] [Medline: [33471111](https://pubmed.ncbi.nlm.nih.gov/33471111/)]
59. Moffa G, Catone G, Kuipers J, et al. Using directed acyclic graphs in epidemiological research in psychosis: an analysis of the role of bullying in psychosis. *Schizophr Bull*. Oct 21, 2017;43(6):1273-1279. [doi: [10.1093/schbul/sbx013](https://doi.org/10.1093/schbul/sbx013)] [Medline: [28521005](https://pubmed.ncbi.nlm.nih.gov/28521005/)]
60. Iorfino F, Varidel M, Marchant R, et al. The temporal dependencies between social, emotional and physical health factors in young people receiving mental healthcare: a dynamic Bayesian network analysis. *Epidemiol Psychiatr Sci*. Sep 8, 2023;32:e56. [doi: [10.1017/S2045796023000616](https://doi.org/10.1017/S2045796023000616)] [Medline: [37680185](https://pubmed.ncbi.nlm.nih.gov/37680185/)]
61. McNally RJ, Heeren A, Robinaugh DJ. A Bayesian network analysis of posttraumatic stress disorder symptoms in adults reporting childhood sexual abuse. *Eur J Psychotraumatol*. 2017;8(sup3):1341276. [doi: [10.1080/20008198.2017.1341276](https://doi.org/10.1080/20008198.2017.1341276)] [Medline: [29038690](https://pubmed.ncbi.nlm.nih.gov/29038690/)]
62. Norman J, Shahar Y, Kuppermann M, Gold B. Decision-theoretic analysis of prenatal testing strategies. *Stanford Medical Informatics*; 1998. URL: https://www.researchgate.net/publication/2526484_Decision-Theoretic_Analysis_of_Prenatal_Testing_Strategies
63. Kornak J, Lu Y. Bayesian decision analysis for choosing between diagnostic/prognostic prediction procedures. *Stat Interface*. 2011;4(1):27-36. [doi: [10.4310/sii.2011.v4.n1.a4](https://doi.org/10.4310/sii.2011.v4.n1.a4)] [Medline: [23243483](https://pubmed.ncbi.nlm.nih.gov/23243483/)]
64. Parmigiani G. Uncertainty and the value of diagnostic information, with application to axillary lymph node dissection in breast cancer. *Stat Med*. Mar 15, 2004;23(5):843-855. [doi: [10.1002/sim.1623](https://doi.org/10.1002/sim.1623)] [Medline: [14981678](https://pubmed.ncbi.nlm.nih.gov/14981678/)]
65. Spirtes P, Scheines R, Causation GC. Prediction, and Search. Springer; 1993. [doi: [10.1007/978-1-4612-2748-9](https://doi.org/10.1007/978-1-4612-2748-9)]

Abbreviations

AI: artificial intelligence
ATE: average treatment effect
BDT: Bayesian decision-theoretic
BN: Bayesian network
CAIRS: causal artificial intelligence recommendation system
DAG: directed acyclic graph
EU: expected utility
Eu: expected subutility
K6: Kessler-6
RS: recommendation system
SCM: structural causal model

Edited by Andrew Coristine; peer-reviewed by Boluwatife Afolabi, Dominic Oliver, Leela Prasad Gorrepati; submitted 15.01.2025; final revised version received 03.04.2025; accepted 04.04.2025; published 19.06.2025

Please cite as:

Varidel M, An V, Hickie IB, Cripps S, Marchant R, Scott J, Crouse JJ, Poulsen A, O'Dea B, McKenna S, Iorfino F
Causal AI Recommendation System for Digital Mental Health: Bayesian Decision-Theoretic Analysis
J Med Internet Res 2025;27:e71305

URL: <https://www.jmir.org/2025/1/e71305>

doi: [10.2196/71305](https://doi.org/10.2196/71305)

© Mathew Varidel, Victor An, Ian B Hickie, Sally Cripps, Roman Marchant, Jan Scott, Jacob J Crouse, Adam Poulsen, Bridianne O'Dea, Sarah McKenna, Frank Iorfino. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 19.06.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on <https://www.jmir.org/>, as well as this copyright and license information must be included.