Reinforcement Learning on AYA Dyads to Enhance Medication Adherence

Ziping Xu[1], Hinal Jajal[1], Sung Won Choi[2], Inbal Nahum-Shani[2], Guy Shani[2], Alexandra M. Psihogios[3], Pei-Yao Hung[2], and Susan Murphy[1]

[1]Harvard University, Cambridge, MA, USA
[2]University of Michigan, Ann Arbor, MI, USA
[3]Northwestern University, Feinberg School of Medicine, Chicago, IL, USA

Abstract. Medication adherence is critical for the recovery of adolescents and young adults (AYAs) who have undergone hematopoietic cell transplantation (HCT). However, maintaining adherence is challenging for AYAs after hospital discharge, who experience both individual (e.g. physical and emotional symptoms) and interpersonal barriers (e.g., relational difficulties with their care partner, who is often involved in medication management). To optimize the effectiveness of a three-component digital intervention targeting both members of the dyad as well as their relationship, we propose a novel Multi-Agent Reinforcement Learning (MARL) approach to personalize the delivery of interventions. By incorporating the domain knowledge, the MARL framework, where each agent is responsible for the delivery of one intervention component, allows for faster learning compared with a flattened agent. Evaluation using a dyadic simulator environment, based on real clinical data, shows a significant improvement in medication adherence (approximately 3%) compared to purely random intervention delivery. The effectiveness of this approach will be further evaluated in an upcoming trial.

Keywords: Reinforcement Learning · Dyadic Relationships · Medication Adherence · Digital Health

1 Introduction

For patients who have undergone allogeneic **hematopoietic stem cell transplantation (HCT)**, strict adherence to medication regimens, such as prophylactic immunosuppressant therapy (i.e., calcineurin inhibitors, such as tacrolimus or cyclosporine, taken twice-daily), is crucial for mitigating the risk of acute graft-versus host disease (GVHD) [4]. Acute GVHD occurs in 50-70% of patients following HCT. A lower medication adherence (60%) rate is shown to associate with higher severity of GVHD [5].

The challenges of adherence management are amplified among **adolescents and young adults (AYAs)**, who often demonstrate poorer medication adherence [15,12,8]. For AYAs with cancer, self-management rarely involves the individual alone. Instead, up to 73% of family care partners bear the primary responsibility for managing cancer-related medications for AYAs [14].

Many of these dyads express a desire to move toward sharing these responsibilities with each other [14]. Indeed, for AYAs with chronic health conditions, this developmental period often marks a shift from relying solely on a caregiver to taking more personal responsibility for health care. While shifts in autonomy versus dependence and navigating the ensuing family conflict that can arise from these new dynamics are normative parts of AYA development, difficult family interactions can have a detrimental impact on medication adherence. For example, in a meta-analysis [13], higher level of family conflict and lower levels of family cohesion were significantly associated with worse medication adherence across pediatric illnesses and age groups.

After being discharged from the hospital, both individuals in the dyad face significant emotional and physical challenges as they adjust to managing medication regimens *outside the hospital environment*. For AYAs, the daily challenges of managing complex medication regimens, coping with treatment side effects, coping with stress, and maintaining normal activities in the context of a complex medical regimen can create distress in their home environment. Similarly, care partners must balance caregiving responsibilities with their personal obligations. Those who shoulder heavy caregiving responsibilities at home face higher physical and emotional stressors, which can impede their ability to provide effective care, make sound decisions, and support their AYA's self-care [16].

This need for support outside the inpatient environment motivates the development of interventions that leverage **digital technologies** such as mobile devices [21]. Digital interventions are promising for supporting both AYAs and care partners *at home* on a daily basis, compared to traditional clinical support delivered with limited frequency (e.g., weekly clinical visits for post-HCT AYAs). There is strong heterogeneity across dyads and the users' context are constantly changing, which makes it important to personalize the intervention delivery to optimize the effectiveness of digital interventions. Reinforcement Learning (RL), a machine learning technique that adaptively learns the optimal behavior in an unknown environment to maximize cumulative rewards, is a promising approach for achieving this personalization. RL has been successfully applied in a variety of digital interventions [7,1,18,3].

In this paper, we describe our work in developing an RL algorithm for ADAPTS-HCT [17]. ADAPTS-HCT is a digital intervention for improving medication adherence by AYAs over 100 days after receiving HCT. ADAPTS-HCT integrates three components: (1) twice-daily messages promoting positive emotions for the AYA, (2) daily messages focusing on coping and self-care strategies for the care partner, and (3) a weekly collaborative game for improving their relationship [17]. *We call the three components AYA, care partner, and relationship component*, respectively. Table 1 summarizes these components. The fully developed intervention package will be evaluated in the upcoming clinical trial.

Goals. Our goal is to design an RL algorithm that can personalize the delivery of these interventions to optimize their effectiveness. Given the complexity of the dyadic structure, we identify the following two key challenges:

Table 1: Intervention components in ADAPTS-HCT

Component	Intervention
AYA	Twice-daily positive psychology messages
Care partner	Daily positive psychology messages
Relationship	Weekly collaborative game designed to facilitate positive dyadic interpersonal relationship

1. **Managing multiple intervention decisions across different multi-scales.** There are three intervention components, each requiring decisions to be made at a different time scales. The decision-making occurs twice daily for AYAs, daily for care partners, and weekly for the relationship component. Making decisions on multiple timescales complicates the algorithm design.
2. **Accelerating learning in noisy, data-limited settings.** Observed data in digital intervention deployment is quite noisy [19]. Furthermore, limited data will be available to support in decision making for dyads recruited early in the clinical trial. Additionally, less data is available for learning decisions that occur at slower timescales. These factors necessitate a sample-efficient algorithm that learns faster given limited data.

Contribution. Our contribution is a novel multi-agent RL (MARL) framework involving three RL agents, where each agent is responsible for making decisions for one specific intervention component and operates at the timescale corresponding to its intervention component timescale, which directly addresses challenge (1) about multi-scale decision-making.

The use of MARL decouples the decision processes of different intervention components, thus improving interpretability of the agent model design. This improved interpretability allows us to incorporate domain knowledge into the agent-specific algorithm designs to address challenge (2). To further accelerate learning, we propose a novel **reward engineering** method that learns a less noisy surrogate reward function for each component. Through evaluation in a carefully designed dyadic environment, we demonstrate both the superior performance of our proposed algorithm and strong collaborative behavior among the three agents. Lack of collaboration is often a critical issue in MARL [9].

2 RL Framework and Domain Knowledge

We start with formulating the intervention decision making as an RL problem, where we underscore the challenge in the multiple time scales. HCT treatment is followed by an outpatient 14-weeks twice-daily medication regimen. Decision times within the 14 weeks are denoted by (w, d, t) where $w \in \{1, \ldots, 14\}$ is the week index, $d \in \{1, \ldots, 7\}$ is the day index, and $t \in \{1, 2\}$ is the decision window within a day.

Primary goal. The primary goal is to make decisions at each decision time t to maximize cumulative sum of medication adherence $\sum_{w=1}^{14} \sum_{d=1}^{7} \sum_{t=1}^{2} R^{\texttt{AYA}}_{w,d,t}$,

Table 2: Summary of variables about each target component

Target	Variable	Type	Description
AYA	$R^{\texttt{AYA}}_{w,d,t}$	binary	Medication adherence at time t on day d in week w
	$A^{\texttt{AYA}}_{w,d,t}$	binary	Intervention at time t on day d in week w
	$B^{\texttt{AYA}}_{w,d,t}$	continuous	App burden at time t on day d in week w
Care partner	$Y^{\texttt{CARE}}_{w,d}$	continuous	Psychological distress on day d in week w
	$A^{\texttt{CARE}}_{w,d}$	binary	Intervention on day d in week w
	$B^{\texttt{CARE}}_{w,d}$	continuous	App burden on day d in week w
Relationship	$Y^{\texttt{REL}}_{w}$	binary	Relationship quality at the end of week w
	$A^{\texttt{REL}}_{w}$	binary	Game intervention at the beginning of week w

where $R^{\texttt{AYA}}_{w,d,t}$ is medication adherence at window t on day d in week w. See Table 2, for selected information that will be collected on the dyad.

Action space. All actions are binary (deliver versus do not deliver intervention content); see Tables 1,2. When the current time ($d = 1, t = 1$) is the first decision time on the first day of the week, the agent chooses a three-dimensional action corresponding to all three interventions components. If the current time is the first time on a day after the first day of the week ($d > 1, t = 1$), the agent chooses a two-dimensional action corresponding to only the AYA intervention and the care partner components. At the second time on each day ($t = 2$) the agent chooses a one-dimensional action corresponding to only the AYA intervention component.

Observation space. Apart from the dynamic action space, we collect observations about different components at different time scales as well; see Table 2. At each time (w, d, t), we collect the current medication adherence and digital intervention burden from the AYA component. In the end of each day d, we collect the psychological distress and digital intervention burden from the care partner component. In the end of a week w, we collect the relationship quality from questionnaires from both the AYA and the care partner.

3 Domain Knowledge through Causal Diagram

Our algorithm design is guided by domain knowledge encoded as the causal diagram in Fig. 1. This diagram describes the scientific team's understanding of the primary causal relationships between the variables in each component listed in Table 2. Note that the causal relationships are likely more complex and direct paths may exist between any two variables. However the scientific team believes that these other paths are likely to be less detectable given the noise in digital intervention data. We summarize the primary pathways that interventions can take to effect the AYA's adherence in the following.

1. **AYA intervention.** The AYA interventions $A^{\texttt{AYA}}_{w,d,t}$ should directly influence the immediate AYA's adherence $R^{\texttt{AYA}}_{w,d,t}$ (black arrows).

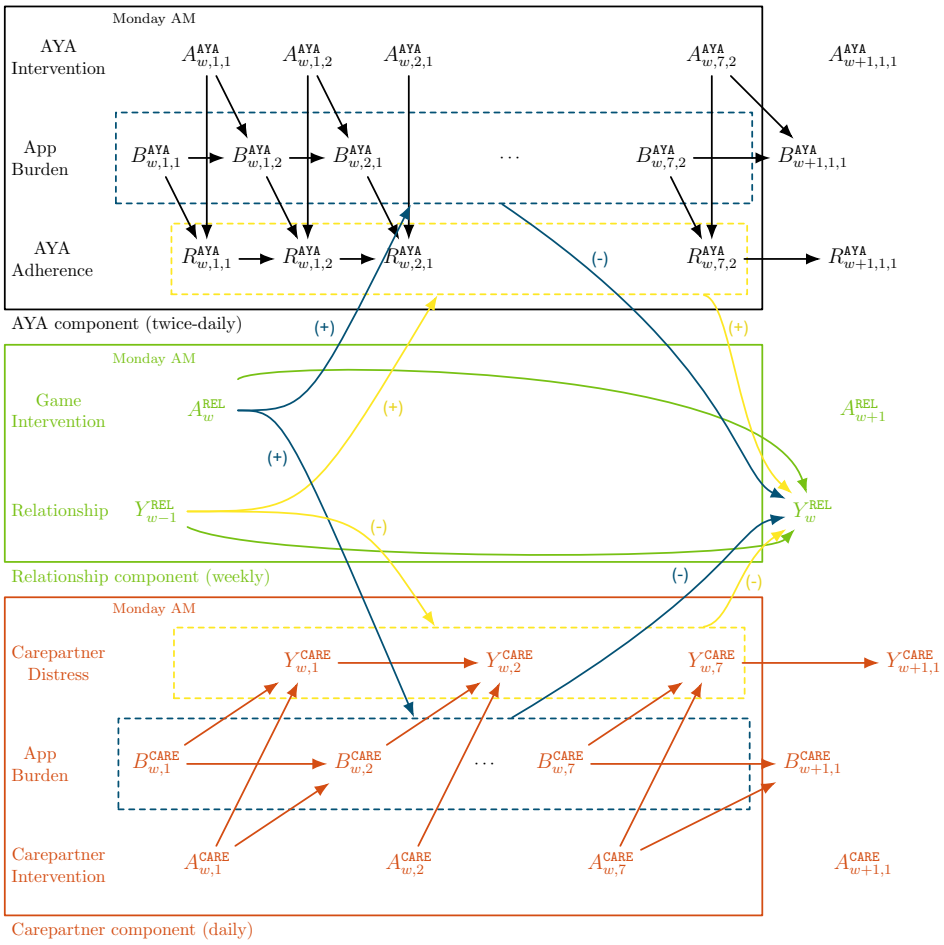


Fig. 1: Causal diagram for ADAPTS-HCT intervention [1]. We categorize the variables into three components: AYA component (marked in **black**), care partner component (marked in red), and relationship component (marked in green). Each component operates at different time scales. Variables in the AYA component evolve on a twice-daily basis, while the care partner component operates on a daily basis. The relationship component operates on a weekly basis. The arrows indicate the direct causal effects.

[1] In the causal inference literature, this is called a causal Directed Acyclic Graph (DAG), a graphical representation of causal relationships among a set of variables [11].

2. **Game intervention.** The game intervention $A_w^{\texttt{REL}}$ has two pathways by which it is expected to effect AYA's adherence. First, $A_w^{\texttt{REL}}$ is expected to increase the AYA's burden $B_{w,d,t}^{\texttt{AYA}}$ throughout the week w. And AYA's burden $B_{w,d,t}^{\texttt{AYA}}$ is expected to decrease the AYA's adherence $R_{w,d,t}^{\texttt{AYA}}$ (blue arrows). Second, the game intervention $A_w^{\texttt{REL}}$ is expected to effect next week AYA's adherence $R_{w+1,d,t}^{\texttt{AYA}}$ by improving the end of the week relationship quality $Y_w^{\texttt{REL}}$ (green arrows).
3. **Care partner intervention.** The care-partner intervention $A_{w,d}^{\texttt{CARE}}$ is expected to effect the AYA's adherence indirectly. First, $A_{w,d}^{\texttt{CARE}}$ should decrease the care partner's psychological distress $Y_{w,d}^{\texttt{CARE}}$, which should increase the end of week relationship quality $Y_w^{\texttt{REL}}$ (yellow arrows). Second, $A_{w,d}^{\texttt{CARE}}$ should increase the care partner's burden $B_{w,d}^{\texttt{CARE}}$, which should decrease the the end of week relationship quality $Y_w^{\texttt{REL}}$ (blue arrows).

We further note that the variables from different components are generally independent conditioned on the bottleneck variables, e.g., the relationship quality that blocks all the paths from the care partner variables to the AYA's adherence. This forms the basis of our multi-agent RL design.

4 Proposed Multi-Agent RL Approach

The conditional independence property observed from Fig. 1 motivates us to design a multi-agent RL (MARL) comprising three agents: the AYA agent, the care partner agent, and the relationship agent. Each makes decisions at different time scales for their own component.

The MARL approach allows us to tailor the agent design choices for each agent to optimize the learning speed. Our base RL algorithm for each agent is Randomized Least Square Value Iteration (RLSVI) [10], which has been proven as stable in deployment of mobile health applications [18,3]. Additionally, we use linear models, which helps in discussions of the algorithm and its parameters with domain scientists.

We construct agent-specific features based on Fig. 1. Specifically, the AYA agent's model uses its own variables ($B_{w,d,t}^{\texttt{AYA}}, R_{w,d,t-1}^{\texttt{AYA}}$) and the variables in the relationship component ($Y_{w-1}^{\texttt{REL}}, A_w^{\texttt{REL}}$). Similarly, the care partner agent uses its own variables, as well as the variables in the relationship component. The relationship agent's model uses $Y_{w-1}^{\texttt{REL}}$, and previous weeks' $B_{w-1,7,2}^{\texttt{AYA}}$, $B_{w-1,7}^{\texttt{CARE}}$, as well as a weighted average of AYA adherence and care partner distress in the past week.

4.1 Surrogate Reward Function Design Through Domain Knowledge

Typical MARL [9] with independent learners considers agents making decisions without communication. In our study, the lack of communication is due to the different time scales–the relationship agent that makes decisions in the beginning of a week may not predict the AYA and care partner agents' decisions throughout the week. This may prevent the agents from **collaborating**. For example, the relationship agent may choose to always intervene so as to improve the relationship

quality (the primary goal of the game intervention), which may not be optimal for the AYA's adherence.

Furthermore, the effects of care partner intervention and the game intervention are highly delayed. The game intervention improves end of week relationship quality with a significant delayed effect onto the adherence in the next week. The care partner intervention (positive messages for the care partner) is designed to mitigate the care partner's psychological distress, which only has indirect and delayed effects on the AYA's adherence.

To address the above two issues, we engineer the reward function to account for the delayed effects and across-component effects of each intervention component to promot collaboration. Similar reward engineering in the context of digital interventions is discussed in [20]. Our approach is distinct in that we explore the principles for incorporating domain knowledge to guide the reward function design.

Domain knowledge informed surrogate reward functions. We introduce the surrogate reward functions for the relationship agent and the care partner agent. As informed by Fig. 1, the delayed effect of the game intervention is through the relationship quality and the AYA burden. This motivates us to fit a linear model to predict the sum of medication adherence within week w, $\sum_{d=1}^{7}\sum_{t=1}^{2} R^{\texttt{AYA}}_{w,d,t}$, using $(1, Y^{\texttt{REL}}_{w-1}, B^{\texttt{AYA}}_{w,1,1}, A^{\texttt{REL}}_{w}, A^{\texttt{REL}}_{w} \cdot Y^{\texttt{REL}}_{w-1})$ as the covariates. To account for the delayed effect, we engineer the surrogate reward function for the relationship agent as: $r^{\texttt{REL}}_{w} = (1, Y^{\texttt{REL}}_{w-1}, B^{\texttt{AYA}}_{w,1,1}, A^{\texttt{REL}}_{w}, A^{\texttt{REL}}_{w} \cdot Y^{\texttt{REL}}_{w-1})\boldsymbol{\beta}^{\texttt{REL}} + \max_a (1, Y^{\texttt{REL}}_{w}, B^{\texttt{AYA}}_{w+1,1,1}, a, a \cdot Y^{\texttt{REL}}_{w})\boldsymbol{\beta}^{\texttt{REL}}$, where $\boldsymbol{\beta}^{\texttt{REL}} \in \mathbb{R}^5$ are Bayesian linear regression estimates. [1] The above reward yields a two-step greedy policy, which is a good enough approximation for the total sum of the medication adherence. We opt for a simple, linear model here because the bias trade-off is justified by the faster learning and reduction in noise.

The design of the care partner agent is similar. A key observation is that the end of the week relationship quality blocks all the paths from the care partner variables to the AYA's adherence. Thus, we fit a linear model to predict the end of week relationship quality $Y^{\texttt{REL}}_{w+1}$ using $(1, Y^{\texttt{CARE}}_{w,d}, B^{\texttt{CARE}}_{w,d+1}, Y^{\texttt{REL}}_{w-1}, A^{\texttt{CARE}}_{w,d})$ as covariates. The surrogate reward function is: $r^{\texttt{CARE}}_{w,d} = (1, Y^{\texttt{CARE}}_{w,d}, B^{\texttt{CARE}}_{w,d+1}, Y^{\texttt{REL}}_{w-1}, A^{\texttt{CARE}}_{w,d})\boldsymbol{\beta}^{\texttt{CARE}}$, where $\boldsymbol{\beta}^{\texttt{CARE}} \in \mathbb{R}^5$ are Bayesian linear regression estimates.

5 Results

We simulate a *dyadic environment* to evaluate the performance of the proposed framework. The environment design should replicate the noise level and structure that we expect to encounter in the forthcoming ADAPTS-HCT clinical trial.

Our environment is based on Roadmap 2.0 dataset involving 171 dyads, each consisting of a patient undergone HCT (target person) and a care partner. Roadmap 2.0 provides daily positive psychology interventions to the care partner

[1] We choose the prior mean to reflect our guesses on the sign the coefficients. The prior variance is chosen so the prior mean dominates until around the 5th dyads. The complete prior is provided in Appendix.

only. Roadmap 2.0 collects wearable devices data, for example, physical activity, and self-report data, for example, mood score.

We build upon the environment design in [6], which also uses the Roadmap 2.0 data, but primarily focuses on AYA and relationship intervention component. We extend the environment to include the care partner intervention component. Specifically, we fit a separate multi-variate linear model for each component's outcome (ie., $R^{\texttt{AYA}}_{w,d,t}, Y^{\texttt{CARE}}_{w,d}, Y^{\texttt{REL}}_{w}$) in the dataset. These models simulate the user trajectories under no intervention.

To simulate outcomes under treatments, we impute the treatment effects of the interventions and the effects of app burden, so the induced standard treatment effects (STE) [2] are around 0.15, 0.3, and 0.5. These STEs are commonly seen in behavioral science studies [2]. A complete description and code of the dyadic environment is provided in supplementary material [3].

5.1 Cumulative Adherence Improvement

We simulate 25 dyads, the planned sample size in the upcoming pilot study, by sampling dyads sequentially with replacement from Roadmap 2.0 dataset. Each dyad is simulated for 14 weeks. We implement the following three algorithms: `SingleAgent`, `MultiAgent`, and `MultiAgent+SurrogateRwd`. Here the `SingleAgent` is the algorithm that trains a single agent that outputs all the three types of actions. The `MultiAgent` is the proposed MARL algorithm using the adherence as the learning reward signal for all three agents. The `MultiAgent+SurrogateRwd` is the proposed MARL algorithm using the surrogate reward functions. The full details of the algorithms are described in supplementary material.

We compare the proposed RL algorithms with a random policy, where $P(A^{\texttt{AYA}}_{w,d,t} = 1) = P(A^{\texttt{CARE}}_{w,d} = 1) = P(A^{\texttt{REL}}_{w} = 1) \equiv 0.5$ in terms of the cumulative adherence improvement.

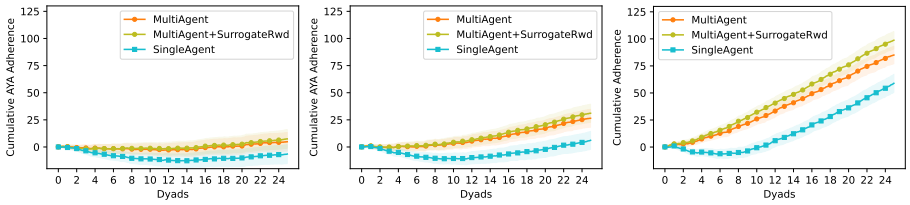
We also observe that all the algorithms can make more significant improvement over the random policy under a higher STE. `SingleAgent` takes longer to learn due to the larger number of parameters compared to `MultiAgent`. We also see an advantage of using surrogate rewards through an increased cumulative adherence at all levels of STE. Notice that for a low STE, the learning is slow, which is intuitive given that the the signal-to-noise ratio is low in such an environment. Additional ablation studies and analysis on the collaborating behavior is provided in the supplementary material.

6 Discussion

In this paper, we propose an MARL algorithm that effectively learns to optimize delivery of the ADAPTS-HCT digital interventions. While this presents a

[2] STE here is defined as the difference in the mean of the primary outcomes under the proposed intervention package and these under no intervention, which is further standardized by the standard deviation under no intervention.

[3] https://github.com/StatisticalReinforcementLearningLab/ADAPTS-HCT-AIME



(a) STE 0.15 (b) STE 0.3 (c) STE 0.5

Fig. 2: Cumulative adherence improvement over the uniform random policy for all three components under dyadic environments with different STEs. The confidence interval is the standard deviation based on 1000 independent runs.

significant step towards preparing for the ADAPTS-HCT clinical trial, several challenges remain to be addressed. First, in the real-clinical trial, the participants are recruited incrementally with significant overlaps, whereas our dyadic environment assumes a simple sequential recruitment. Second, the clinical trial study emphasizes the need for after-study analysis, such as causal inference on treatment effects, which often requires smooth allocation functions [22]. Additionally, there is room to further improve algorithm performance. For example, our proposed algorithm pools data across dyads to reduce learning variance but does not account for heterogeneity across dyads. The algorithm may benefit from a more flexible pooling, e.g., a random effect model.

References

1. Samuel L Battalio, David E Conroy, Walter Dempsey, and et al. Sense2stop: a micro-randomized trial using wearable sensors to optimize a just-in-time-adaptive stress management intervention for smoking relapse prevention. *Contemporary Clinical Trials*, 109:106534, 2021.
2. Jacob Cohen. *Statistical power analysis for the behavioral sciences*. routledge, 2013.
3. Susobhan Ghosh, Yongyi Guo, Pei-Yao Hung, Lara Coughlin, Erin Bonar, Inbal Nahum-Shani, Maureen Walton, and Susan Murphy. Miwaves reinforcement learning algorithm. *arXiv preprint arXiv:2408.15076*, 2024.
4. B Gresch, M Kirsch, K Fierz, JP Halter, G Nair, K Denhaerynck, and S De Geest. Medication nonadherence to immunosuppressants after adult allogeneic haematopoietic stem cell transplantation: a multicentre cross-sectional study. *Bone marrow transplantation*, 52(2):304–306, 2017.
5. M Kirsch, A Gotz, JP Halter, U Schanz, G Stussi, F Dobbels, and S De Geest. Differences in health behaviour between recipients of allogeneic haematopoietic sct and the general population: a matched control study. *Bone marrow transplantation*, 49(9):1223–1230, 2014.
6. Shuangning Li, Lluis Salvat Niell, Sung Won Choi, Inbal Nahum-Shani, Guy Shani, and Susan Murphy. Dyadic reinforcement learning. *arXiv preprint arXiv:2308.07843*, 2023.
7. Peng Liao, Kristjan Greenewald, Predrag Klasnja, and Susan Murphy. Personalized heartsteps: A reinforcement learning algorithm for optimizing physical activity. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 4(1):1–22, 2020.

8. Karen S Lyons and Christopher S Lee. The theory of dyadic illness management. *Journal of family nursing*, 24(1):8–28, 2018.
9. Afshin Oroojlooy and Davood Hajinezhad. A review of cooperative multi-agent deep reinforcement learning. *Applied Intelligence*, 53(11):13677–13722, 2023.
10. Ian Osband, Benjamin Van Roy, and Zheng Wen. Generalization and exploration via randomized value functions. In *International Conference on Machine Learning*, pages 2377–2386. PMLR, 2016.
11. Judea Pearl et al. Models, reasoning and inference. *Cambridge, UK: CambridgeUniversityPress*, 19(2):3, 2000.
12. Alexandra M Psihogios, Annisa M Ahmed, Elise R McKelvey, Toto, and et al. Social media to promote treatment adherence among adolescents and young adults with chronic health conditions: A topical review and tiktok application. *Clinical Practice in Pediatric Psychology*, 10(4):440, 2022.
13. Alexandra M Psihogios, Heather Fellmeth, Lisa A Schwartz, and Lamia P Barakat. Family functioning and medical adherence across children and adolescents with chronic health conditions: a meta-analysis. *Journal of Pediatric Psychology*, 44(1):84–97, 2019.
14. Alexandra M Psihogios, Lisa A Schwartz, Kylie B Ewing, Bryn Czerniecki, Leslie S Kersun, Ahna LH Pai, Janet A Deatrick, and Lamia P Barakat. Adherence to multiple treatment recommendations in adolescents and young adults with cancer: a mixed methods, multi-informant investigation. *Journal of adolescent and young adult oncology*, 9(6):651–661, 2020.
15. Alexandra M Psihogios, Colleen Stiles-Shields, and Martha Neary. The needle in the haystack: identifying credible mobile health apps for pediatric populations during a pandemic and beyond. *Journal of Pediatric Psychology*, 45(10):1106–1113, 2020.
16. Susan C Reinhard, Barbara Given, Nirvana Huhtala Petlick, and Ann Bemis. Supporting family caregivers in providing care. *Patient safety and quality: An evidence-based handbook for nurses*, 2008.
17. Guy Shani, Sung Won Choi, Susan Murphy, Inbal Billie Nahum-Shani, and et al. Designing a dyadic just-in-time adaptive intervention for medication adherence post-hematopoietic cell transplantation. In *8th Annual Technology in Psychiatry Summit (TIPS)*, Phoenix, AZ, December 2024.
18. Anna L Trella, Kelly W Zhang, Hinal Jajal, Inbal Nahum-Shani, Vivek Shetty, Finale Doshi-Velez, and Susan A Murphy. A deployed online reinforcement learning algorithm in an oral health clinical trial. *arXiv preprint arXiv:2409.02069*, 2024.
19. Anna L Trella, Kelly W Zhang, Inbal Nahum-Shani, Vivek Shetty, Finale Doshi-Velez, and Susan A Murphy. Designing reinforcement learning algorithms for digital interventions: pre-implementation guidelines. *Algorithms*, 15(8):255, 2022.
20. Anna L Trella, Kelly W Zhang, Inbal Nahum-Shani, Vivek Shetty, Finale Doshi-Velez, and Susan A Murphy. Reward design for an online reinforcement learning algorithm supporting oral self-care. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 15724–15730, 2023.
21. Fabio Alexis Rincón Uribe, Maria Fernanda Monteiro Favacho, Paula Marília Nascimento Moura, and et al. Effectiveness of an app-based intervention to improve well-being through cultivating positive thinking and positive emotions in an adult sample: study protocol for a randomized controlled trial. *Frontiers in Psychology*, 14:1200960, 2023.
22. Kelly W Zhang, Nowell Closser, Anna L Trella, and Susan A Murphy. Replicable bandits for digital health interventions. *arXiv preprint arXiv:2407.15377*, 2024.