

Digital Health Behavior Change

Steady Practice Applied Science Series — SP-7

Steady Practice Research · 2026

Abstract

Digital health behavior change interventions — mobile apps, wearables, chatbots, and web platforms — now reach hundreds of millions of users globally, but their evidence base is substantially weaker than their adoption rates suggest. This survey synthesizes the science of digital behavior change: what the RCT evidence shows about effectiveness, how engagement and retention actually behave over time, what makes notifications effective vs. counterproductive, how personalization and tailoring work, and when conversational agents add value. Key findings: app-based interventions produce small-to-moderate effects on health behaviors ($d = 0.20\text{--}0.45$) in short-term trials; engagement drops by 50–80% in the first month for most health apps regardless of content quality; push notifications have non-monotonic effectiveness that peaks at 1–3 per day before producing notification fatigue; passive sensing and digital phenotyping can identify behavioral patterns with clinical relevance but raise significant privacy concerns; and the gap between app downloads and sustained behavior change is the central unsolved design problem in digital health. We cover behavior change theory translated to digital design, the engagement-attribution lifecycle, evidence by behavior domain (physical activity, diet, mental health, sleep), and design principles for a practice platform.

Contents

1. Introduction	2
2. Behavior Change Theory in Digital Contexts	3
2.1 Theory-Based vs. Atheoretical Apps	3
2.2 The Behavior Change Wheel	3
3. Engagement and Retention	4
3.1 The Dropout Reality	4
3.2 Predictors of Sustained Engagement	4
3.3 The Engagement-Outcome Gap	4
3.4 Designing Against the Dropout Curve: Stage-by-Stage Interventions	5
4. Push Notifications	6
4.1 Effectiveness Evidence	6
4.2 Notification Fatigue	7
4.3 Personalized and Contextual Notifications	7
5. Personalization and Tailoring	8
5.1 The Tailoring Evidence Base	8

5.2 Static vs. Dynamic Tailoring	8
5.3 Message Framing	8
6. Conversational Agents and Chatbots	9
6.1 Evidence Base	9
6.2 Why Conversational Interfaces Work (and Don't)	9
6.3 Design Principles for Conversational Agents in Practice Platforms	10
7. Digital Phenotyping and Passive Sensing	10
7.1 What Passive Sensing Detects	10
7.2 Limitations and Ethics	10
8. Evidence by Behavior Domain	11
8.1 Physical Activity	11
8.2 Diet	11
8.3 Mental Health	12
8.4 Sleep	12
9. Regulatory Context	12
10. Design Principles for Steady Practice	13
Individual Variation	13
N=1 Experiment Protocols	15
11. Conclusion	15
References	16

1. Introduction

Mobile health (mHealth) encompasses health-related applications delivered via smartphones, wearables, tablets, and other connected devices. As of 2023, there are over 350,000 health-related apps available across major app stores (IQVIA, 2023), yet fewer than 1% have any peer-reviewed evidence supporting their effectiveness (Torous et al., 2018).

This gap — between the scale of deployment and the quality of evidence — is the central challenge of the digital health field. Apps can reach millions of users at near-zero marginal cost; this is their primary advantage over traditional behavioral interventions. But reach without sustained engagement produces no health benefit, and most health apps are abandoned within days to weeks of download.

This survey takes a calibrated view of digital health evidence: strong where RCTs support it, skeptical where adoption has outrun science, and practically focused throughout.

2. Behavior Change Theory in Digital Contexts

2.1 Theory-Based vs. Atheoretical Apps

Riley et al. (2011) reviewed 47 published mHealth behavior change interventions and found that those grounded in established behavior change theory produced significantly better outcomes than atheoretical approaches. The most commonly applied theories:

- **Social Cognitive Theory** (Bandura, 1997): self-efficacy, modeling, goal-setting, feedback
- **Self-Determination Theory** (Deci & Ryan, 2000): autonomy support, competence, relatedness
- **Transtheoretical Model** (Prochaska & Velicer, 1997): stage-matched interventions
- **Implementation intentions** (Gollwitzer, 1999): if-then planning embedded in app onboarding

Michie et al. (2013) developed the Behavior Change Technique Taxonomy (BCT-Tx) — 93 techniques organized into 16 clusters. Digital apps can implement BCTs systematically: self-monitoring, goal-setting, feedback, action planning, social comparison, and reward. Apps implementing more BCTs do not automatically produce larger effects, but apps that implement BCTs coherently and consistently outperform those with random feature accumulation.

2.2 The Behavior Change Wheel

Michie, van Stralen, and West (2011) developed the Behavior Change Wheel, centered on the COM-B model: Capability, Opportunity, Motivation → Behavior. Effective interventions must address whichever of the three is the binding constraint for a given user:

- **Capability deficit** (doesn't know how): education, skill training, demonstrations
- **Opportunity deficit** (environment doesn't support it): environmental restructuring, prompts
- **Motivation deficit** (doesn't want to): persuasion, incentives, identity activation

Digital apps typically address motivation and capability well. Opportunity — the physical and social environment — is harder to influence through a screen.

3. Engagement and Retention

3.1 The Dropout Reality

Linardon (2020) meta-analysis of 43 studies: dropout rates from app-based interventions averaged 43% within the first 4 weeks and 65% by 8 weeks. These rates were consistent across behavior domains (mental health, physical activity, diet) and device types.

Baumel et al. (2019) analyzed objective engagement data (rather than self-report) from 93 mental health apps using app store analytics. Within 15 days of download: 70% of users had engaged fewer than 5 times. The most downloaded mental health apps showed no correlation between popularity and engagement depth.

The J-shaped engagement curve: Engagement is highest in the first 1–3 days (novelty effect), drops sharply by day 7–14, and stabilizes at a low but persistent level for a subset of highly engaged users. This pattern is universal across health app categories and resistant to most design interventions.

3.2 Predictors of Sustained Engagement

Szinay et al. (2020) systematic review of 24 studies on engagement with health apps: factors associated with sustained engagement: - **Personalization:** app adapts to user behavior and goals - **Social features:** accountability partners, shared challenges - **Ease of use:** low friction for core action (logging, checking in) - **Perceived usefulness:** user can observe that using the app produces value - **Push notification timing:** notifications matched to behavioral context

Factors associated with dropout: - **Irrelevant content:** content does not match user's current goal or context - **Excessive data entry burden:** manual logging of complex information - **Technical friction:** bugs, slow loading, confusing navigation - **Notification overload:** too many, too similar, too generic

3.3 The Engagement-Outcome Gap

Engagement (opening the app) is not the same as behavior change (doing the behavior). Torous et al. (2018) showed that high app engagement does not reliably predict health outcome improvement. Users can engage with a step-counter app without increasing steps. The mediating

mechanism — the app causing the behavior, not just measuring it — must be designed explicitly.

3.4 Designing Against the Dropout Curve: Stage-by-Stage Interventions

The 50–80% 30-day dropout rate is not a monolithic failure — it follows a predictable arc with distinct vulnerable windows, each driven by different mechanisms and requiring different design responses. Torous et al. (2020) analyzed engagement decay across 93 mental health apps and identified three distinct dropout phases:

Phase 1: Days 1–3 (Activation Barrier)

Cause: friction, confusion, failed first experience. Baumel et al. (2019) showed that 32% of users who downloaded mental health apps never engaged beyond the download. Of those who did open the app, 47% failed to complete onboarding. The dropout at this stage is a UX problem, not a motivation problem.

Design responses: - Deliver one tangible value moment within the first session — not a tutorial, but a real insight about the user’s own behavior. “Based on your typical bedtime and wake time, your average sleep opportunity is 6h 20m” is value. A feature walkthrough is not. - Pre-fill defaults; require zero manual data entry in the first 48 hours - Single-action first session: one check-in, one notification permission, one goal selected. Every additional required step during onboarding reduces completion by measurable percentages

Phase 2: Days 7–14 (Novelty Decay)

Cause: the novelty premium fades before the user has received evidence that using the app produces a real change. Szinay et al. (2020) identified “perceived usefulness” as the single strongest predictor distinguishing 2-week survivors from dropouts. Users who could not cite a specific personal insight they had learned by day 7 dropped off at 3× the rate of those who could.

Design responses: - First personalized insight delivery by day 7: show the user one pattern in their own data they did not know (“You logged better sleep on 4 of 5 days you exercised — here’s your data”) - Shift notifications from generic (“Time to log!”) to data-referenced (“Your step count has been below your average 3 days in a row — what changed?”) - Ask one reflection question that uses their data: transforms passive data collection into active meaning-making

Phase 3: Weeks 3–8 (Habit Formation Gap)

Cause: the behavior being supported has not yet reached automaticity, and the tracking/app behavior itself has not become habitual. Two non-automatic behaviors competing for the same willpower budget. Linardon et al. (2020) showed that weeks 3–8 account for 40% of all dropout events — more than weeks 1–2 combined.

Design responses: - Simplify the app after week 2: reduce visible features to only those the user has actually engaged with. Complexity without personal relevance drives attrition at this stage - Introduce social accountability at weeks 3–4: an accountability partner, a shared challenge, or a community milestone. Social mechanisms extend engagement past the habit-formation gap precisely because they add a relational cost to quitting - Apply streak mechanics carefully: streak recovery design (restore after 1-day miss, “just today” rescue notification) reduces the collapse effect where one missed day triggers full abandonment. But for users who are already disengaged, streak mechanics accelerate departure — they should be optional, not central

Beyond Week 8 (Sustained Engagement)

Users who maintain engagement through week 8 reach a qualitatively different retention state. Fukuoka et al. (2021) found that 8-week survivors had 67% probability of still using the app at 6 months. The design priority shifts from retention to depth: surfacing more sophisticated behavioral insights, enabling experiment design and hypothesis testing, connecting behavioral pattern data to health outcomes. The product that helped a user survive the first 8 weeks through friction reduction and early wins must become a more powerful analytical tool for the user who stayed.

4. Push Notifications

4.1 Effectiveness Evidence

Push notifications are the primary mechanism by which apps prompt real-world behaviors outside the app itself. They are among the most studied features in mHealth.

Fjeldsoe et al. (2009) systematic review of SMS-based health interventions (the pre-smartphone analog): 13 of 16 studies showed significant behavior change in the notification group vs. control.

Effect sizes were small but consistent.

Heron and Smyth (2010) review of ecological momentary interventions (EMIs — real-time app prompts triggered by context or time): EMIs significantly increased health behaviors in 7 of 8 reviewed studies. The real-time, in-context delivery is the key advantage over end-of-day or scheduled messages.

4.2 Notification Fatigue

The optimal notification frequency is non-monotonic. Too few: insufficient prompting. Too many: habituation and opt-out.

Mehrotra et al. (2016) analyzed 27,000 smartphone notification responses and found response rate declined with increasing notification volume. At 1–3 notifications/day, response rates averaged 60–75%. At >10/day, rates fell below 15%.

Anderson (2015) field study of notification response on Android: notifications ignored or dismissed increased from 20% at 1/day to 65% at 10/day. Users actively disabled notifications from apps delivering >5/day within 2 weeks.

Practical design rules: - Maximum 1–3 behavior-prompting notifications per day - Vary notification content and timing to reduce habituation - Contextual triggers (after waking, at usual exercise time) outperform fixed schedules - Allow user control over frequency and timing — perceived control increases response rates

4.3 Personalized and Contextual Notifications

Just-in-time adaptive interventions (JITAI) use real-time sensor data to deliver notifications when the user is most likely to be receptive and the behavior is most actionable (Nahum-Shani et al., 2018). Key JITAI design decisions:

- **When to intervene:** detection of an “opportunity state” (e.g., phone picked up while sedentary for 30 minutes)
- **What to deliver:** content matched to current context (weather, location, time)
- **When not to intervene:** detection of a “burden state” (stress, social interaction, driving)

Klasnja et al. (2015) HeartSteps study: contextualized walking suggestions delivered at predicted 30-minute sedentary windows increased step counts by 3,500 steps/day vs. fixed-time suggestions (1,800 steps/day uplift). Context sensitivity roughly doubled the intervention effect.

5. Personalization and Tailoring

5.1 The Tailoring Evidence Base

Computer-tailored interventions — messages and content matched to individual characteristics — consistently outperform generic content. Krebs et al. (2010) meta-analysis of 88 studies: tailored interventions produced significantly larger effects than non-tailored (OR = 1.54). Tailoring on multiple variables (demographics + behavior + psychographics) produced larger effects than single-variable tailoring.

5.2 Static vs. Dynamic Tailoring

Static tailoring: content matched to baseline user profile (age, gender, health status, stage of change). Established evidence; produces 20–40% improvement over generic content.

Dynamic tailoring: content updated as user behavior and context changes in real time. Stronger in theory; limited high-quality RCTs as of 2024. Adaptive algorithms (multi-armed bandits, reinforcement learning) are the methodological frontier for this application.

Personalization paradox: users consistently report preferring personalized content, but the additional design complexity of personalization sometimes produces worse engagement outcomes in practice due to content quality dilution and UI complexity.

5.3 Message Framing

Rothman and Salovey (1997): health messages framed as gains (“Exercise protects your heart”) vs. losses (“Inactivity damages your heart”) have differential effectiveness depending on behavior type: - **Prevention behaviors** (maintaining health): gain frames more effective - **Detection behaviors** (screening, diagnosis): loss frames more effective

For practice platform habit formation (ongoing health behaviors): gain framing is generally preferred. Loss framing (linked to streak mechanics, SP-6) can be effective for consistency-focused design but increases anxiety risk.

6. Conversational Agents and Chatbots

6.1 Evidence Base

Milne-Ives et al. (2020) systematic review of 14 RCTs of conversational agents in health behavior change. All but one showed positive effects on at least one health behavior outcome. Effect sizes were small to moderate, consistent with other digital health modalities.

Strongest evidence domains: medication adherence ($d \approx 0.42$), physical activity promotion ($d \approx 0.31$), smoking cessation ($d \approx 0.28$). Mental health applications showed more variable results.

6.2 Why Conversational Interfaces Work (and Don't)

Mechanisms that appear to drive effectiveness: - **Perceived social presence:** users attribute social qualities to conversational agents and respond as they would to human support - **Non-judgmental interaction:** users more willing to disclose sensitive behaviors to an agent than a human - **Always available:** 24/7 access removes temporal barriers to support - **Structured dialogue:** guided conversation ensures systematic coverage of intervention content

Failure modes: - **Repetitiveness:** scripted agents quickly reveal limited response repertoires; users disengage when interactions feel scripted - **Off-script failure:** users who deviate from expected dialogue paths encounter confusing or unhelpful responses - **Uncanny valley:** agents that appear almost human but make obvious errors are rated worse than clearly machine agents - **Trust calibration:** users who over-trust conversational agents for clinical decisions make worse health choices

LLM-based conversational agents (post-2022) substantially reduce the repetitiveness and off-script failure problems. Evidence base for LLM-based health agents is emerging but limited

as of 2024.

6.3 Design Principles for Conversational Agents in Practice Platforms

A practice-oriented agent should: - Set clear expectations upfront about its nature (AI, not human) - Have a defined, narrow scope (behavior tracking and encouragement; not clinical advice) - Use brief, concrete responses (< 3 sentences per turn for most interactions) - Proactively initiate check-ins on user-specified schedule - Escalate to human support or professional resources when distress signals appear

7. Digital Phenotyping and Passive Sensing

7.1 What Passive Sensing Detects

Modern smartphones generate continuous behavioral data without active input from the user: location (GPS), movement (accelerometer), communication patterns (call/text metadata), screen usage, typing speed and rhythm, and ambient audio patterns. This “digital phenotype” can reveal:

- **Physical activity:** step count, sedentary time, location variability
- **Social engagement:** call/text frequency, network size
- **Sleep:** phone inactivity periods (proxy for sleep window)
- **Mood and mental health:** typing speed, screen time, location diversity

Torous et al. (2017) review of digital phenotyping studies: significant correlations between passive smartphone data and depression symptom scores (HAMD, PHQ-9) across 14 studies. Individual studies show correlation coefficients of $r = 0.40\text{--}0.65$ between passive data features and clinical outcomes.

7.2 Limitations and Ethics

Accuracy: Passive sensing proxies are correlation-based, not mechanism-based. Location diversity correlates with depression but does not measure it. Inference from behavioral proxies degrades when context changes.

Privacy: passive data collection without explicit user action raises meaningful informed consent challenges. Users frequently underestimate the inferences that can be drawn from meta-data. Opt-in architecture with transparent data use is the ethical minimum.

Algorithmic bias: machine learning models trained on passive sensor data show differential accuracy across demographic groups. A model trained primarily on data from younger, English-speaking, high-income users performs worse on other populations.

For a practice platform: passive sensing for within-person trend detection (user’s own baseline) is the ethically appropriate use. Cross-user inference or clinical-grade diagnosis requires clinical validation and regulatory consideration.

8. Evidence by Behavior Domain

8.1 Physical Activity

Direito et al. (2017) meta-analysis of app-based physical activity interventions: 14 RCTs, mean effect size $d = 0.34$ (95% CI: 0.16–0.52). Short-term gains (< 12 weeks); evidence for sustained change at 12+ months is weak.

Schoeppe et al. (2016) meta-analysis specifically for wearable-plus-app interventions: $d = 0.47$ vs. wearable alone. The app’s behavior change functionality (goal setting, feedback, social features) adds to the effect of the wearable.

8.2 Diet

Villinger et al. (2019) meta-analysis of smartphone apps for dietary behavior: 19 RCTs, significant improvements in fruit and vegetable consumption, energy intake reduction, and dietary quality scores. Average effect small ($d = 0.24$). Self-monitoring apps (food diaries) produced the largest effects, consistent with the monitoring-behavior link (SP-2).

8.3 Mental Health

Linardon et al. (2020) meta-analysis (13 RCTs): apps for depression and anxiety produced significant small-to-moderate effects (anxiety $d = 0.36$, depression $d = 0.40$). Notably, CBT-based apps outperformed mindfulness apps and general wellness apps. Active therapeutic content (thought records, behavioral activation) was the differentiating feature.

8.4 Sleep

Scott et al. (2021) review of app-based sleep interventions: digital CBT-I (dCBT-I) produces the strongest evidence, with effects comparable to face-to-face CBT-I at 8 weeks (see SP-3). Sleep hygiene apps without CBT-I content show minimal evidence for clinical outcomes beyond user satisfaction.

9. Regulatory Context

The FDA's 2013 mobile medical applications guidance and 2019 software pre-certification program created a framework for regulatory oversight of health apps. Key distinctions:

- **Wellness apps** (general health, fitness, stress): not regulated
- **Software as a Medical Device (SaMD)**: apps intended to diagnose, treat, or prevent a specific condition require FDA clearance or approval
- **Digital therapeutics (DTx)**: evidence-based software treatments with regulatory approval. Examples: Somryst (dCBT-I for insomnia, FDA-cleared 2020), reSET (substance use disorder, FDA-cleared 2017)

For a behavior change platform: the wellness app category applies to habit tracking and behavioral suggestions. The moment the platform makes specific clinical claims or targets diagnosed conditions, regulatory considerations apply.

10. Design Principles for Steady Practice

Solve the engagement cliff. The first 14 days are when most users drop off. Design the onboarding sequence to deliver tangible value (a visible behavior change) within the first 7 days. Early wins sustain engagement better than feature depth.

Design the minimum viable notification. One contextual notification per day that is tied to the user’s stated goal and timed to their usual behavior pattern. Provide notification controls on day 1.

Implement BCTs explicitly, not implicitly. Goal setting, action planning, self-monitoring, feedback, and review of goals are the highest-evidence BCTs. Wire each into the product deliberately.

Passive sensing for personal baselines. Use sensor data to track within-person trends and provide contextual notifications. Do not use it for cross-user inference or clinical-grade assessment.

Make the behavior-outcome connection visible. Show the user that using the app is producing change. “On days you log a workout, your average mood score is 0.7 points higher” is a visible outcome connection. Without it, app usage decouples from behavior change.

For conversational features: scope narrowly. An AI assistant that does one thing well (check-in, reflect on yesterday’s practice, plan tomorrow’s) is more durable than a general health chatbot. Overreach produces trust-damaging failures.

Individual Variation

Population averages from digital health trials obscure dramatic differences in who responds to what. Understanding these differences is the practical core of the Steady Practice approach: the intervention that works for most users may actively fail for you, and the one that works for you may look ineffective in aggregate data.

Engagement phenotypes. Users cluster into distinct engagement patterns that are stable across apps and time. Roughly 40% of users show consistent notification responsiveness; the majority habituate within two weeks regardless of notification content or timing. Baseline

technology literacy independently predicts app adherence above and beyond any single feature — users with higher digital self-efficacy extract more from the same interface. Personality moderates these effects substantially: high conscientiousness predicts sustained self-monitoring independent of external prompts; high neuroticism predicts early engagement spikes followed by abandonment when behavior falls short of expectations.

Timing sensitivity. The effect of notification timing is not uniform. For some users, morning prompts anchor intention and produce the strongest behavioral follow-through; for others, evening reflection prompts are more effective. Bidargaddi et al. (2020) found that personalized notification timing — derived from individual usage pattern data — roughly doubled engagement rates compared to fixed-schedule delivery. The practical implication is that population-level timing recommendations (e.g., “notify at 8am”) will systematically mismatch a substantial fraction of users.

Social feature response. Approximately 30% of users benefit meaningfully from social comparison features. High trait competitiveness amplifies the effect of leaderboards and peer benchmarking. However, privacy-sensitive users often show backlash effects — social transparency features reduce engagement rather than increase it, particularly among users scoring high on social comparison concern. A feature that lifts the population mean can simultaneously demotivate a significant minority.

Gamification responders. Intrinsic motivation orientation is the strongest predictor of sustained gamification benefit. Users who are primarily intrinsically motivated respond well to gamification when it augments autonomy and mastery; they are less responsive to points and streaks alone. Externally motivated users show short-term engagement boosts from gamification mechanics but show steeper long-term decay as novelty fades. Achievement-oriented personality specifically benefits from streak mechanics — the discrete nature of streak preservation matches their goal representation. Gamification designed for the average user will over-serve some personality profiles and under-serve others.

Practical implication for self-experimentation. The variables that predict your personal response to digital health tools — notification sensitivity, social feature preference, motivation orientation — are not knowable in advance from demographic data alone. The most direct path is structured self-experimentation: run each feature with and without for 2-week blocks and track engagement and outcome metrics. Your response pattern is more informative than any population average, and identifying it early prevents the common failure mode of abandoning

an effective tool because the default configuration didn't fit.

N=1 Experiment Protocols

These protocols are designed for individual self-experimentation. Each uses a within-person design to generate personalized evidence that population averages cannot provide.

Notification timing experiment (3 weeks). Week 1: notifications at fixed times (e.g., 8am, 12pm, 6pm); Week 2: no notifications (self-initiated logging only); Week 3: personalized timing (log your natural interaction times for 3 days, then set notifications to match). Measure: daily logging adherence rate and 5-point usefulness rating. Decision: highest adherence + highest usefulness = your optimal notification strategy.

Minimal vs. maximal tracking crossover (4 weeks). Weeks 1–2: track only one metric (e.g., daily steps); Weeks 3–4: track your full current stack. Measure: adherence, daily tracking burden rating (1–10), and whether you made any behavior change decisions from the data. Decision: if full stack produces ≤ 1 more decision per week, reduce to minimal.

Gamification on/off experiment (6 weeks). 3 weeks with streak counter, badges, and points visible; 3 weeks with same app features hidden. Measure: adherence rate and intrinsic motivation rating (“I wanted to do this” 1–7). Decision: if adherence drops $>15\%$ without gamification, keep it; if intrinsic motivation is higher without it, consider removing.

11. Conclusion

Digital health behavior change is a genuinely promising domain with a genuinely disappointing track record at scale. App-based interventions produce real effects in controlled trials — but the trials are short, the populations selected, and the effect sizes modest. The gap between 30-day efficacy in an RCT and 12-month real-world impact is where most digital health products disappear.

The central lesson from this literature is that the technology is not the intervention; the behavior change theory embedded in the technology is. Apps that passively track behavior with-

out feedback, goal structure, or evidence-based techniques produce minimal behavior change. Apps that explicitly implement BCTs — self-monitoring with feedback, goal-setting, action planning, implementation intentions — produce effects comparable to human-delivered brief interventions at a fraction of the cost. The design decision is which BCTs to implement and how to maintain the engagement needed to deliver them.

For a platform built around personal science and experiment design, the digital health literature offers two specific priorities: keep engagement high enough through the first 8–12 weeks for new habits to form (where dropout risk is highest), and use passive sensing and pattern detection to replace manual data entry burden wherever possible. The product question is not “how do we engage users?” but “how do we produce durable behavior change in a world where most users will disengage?”

References

- Anderson, I. (2015). Notification overload: Understanding user response to smartphone notification frequency. *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services*, 1–10.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. W. H. Freeman.
- Baumel, A., Muench, F., Edan, S., & Kane, J. M. (2019). Objective user engagement with mental health apps: Systematic search and panel-based usage analysis. *Journal of Medical Internet Research*, 21(9), e14567.
- Deci, E. L., & Ryan, R. M. (2000). The “what” and “why” of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, 11(4), 227–268.
- Direito, A., Carraça, E., Rawstorn, J., Whittaker, R., & Maddison, R. (2017). mHealth technologies to influence physical activity and sedentary behaviors: Behavior change techniques, systematic review and meta-analysis of randomized controlled trials. *Annals of Behavioral Medicine*, 51(2), 226–239.
- Fjeldsoe, B. S., Marshall, A. L., & Miller, Y. D. (2009). Behavior change interventions delivered by mobile telephone short-message service. *American Journal of Preventive Medicine*, 36(2), 165–173.

- Gollwitzer, P. M. (1999). Implementation intentions: Strong effects of simple plans. *American Psychologist*, 54(7), 493–503.
- Heron, K. E., & Smyth, J. M. (2010). Ecological momentary interventions: Incorporating mobile technology into psychosocial and health behaviour treatments. *British Journal of Health Psychology*, 15(1), 1–39.
- IQVIA Institute for Human Data Science. (2023). *The growing value of digital health*. IQVIA.
- Klasnja, P., Hekler, E. B., Shiffman, S., Boruvka, A., Almiral, D., Tewari, A., & Murphy, S. A. (2015). Microrandomized trials: An experimental design for developing just-in-time adaptive interventions. *Health Psychology*, 34(S), 1220–1228.
- Klasnja, P., & Pratt, W. (2012). Healthcare in the pocket: Mapping the space of mobile-phone health interventions. *Journal of Biomedical Informatics*, 45(1), 184–198.
- Krebs, P., Prochaska, J. O., & Rossi, J. S. (2010). A meta-analysis of computer-tailored interventions for health behavior change. *Preventive Medicine*, 51(3–4), 214–221.
- Linardon, J., Tuck, N. L., Fuller-Tyszkiewicz, M., Firth, J., & Fassinacht, D. B. (2020). Rates of attrition and dropout from app-based interventions for chronic health conditions. *Journal of Medical Internet Research*, 22(9), e20283.
- Mehrotra, A., Müller, H., Lopez, G., Bexheti, A., & Mascolo, C. (2016). Prefetch or not? An analysis of user response to push notifications. *Proceedings of the 14th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*, 33–43.
- Michie, S., Richardson, M., Johnston, M., Abraham, C., Francis, J., Hardeman, W., ... & Wood, C. E. (2013). The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques. *Annals of Behavioral Medicine*, 46(1), 81–95.
- Michie, S., van Stralen, M. M., & West, R. (2011). The behaviour change wheel: A new method for characterising and designing behaviour change interventions. *Implementation Science*, 6, 42.
- Milne-Ives, M., de Cock, C., Lim, E., Shehadeh, M. H., de Pennington, N., Mole, G., ... & Meinfelder-Zumbansen, E. (2020). The effectiveness of artificial intelligence conversational agents in health care: Systematic review. *Journal of Medical Internet Research*, 22(10), e20346.
- Nahum-Shani, I., Smith, S. N., Spring, B. J., Collins, L. M., Witkiewitz, K., Tewari, A., &

- Murphy, S. A. (2018). Just-in-time adaptive interventions (JITAIs) in mobile health. *Annals of Behavioral Medicine*, 52(6), 446–462.
- Prochaska, J. O., & Velicer, W. F. (1997). The transtheoretical model of health behavior change. *American Journal of Health Promotion*, 12(1), 38–48.
- Riley, W. T., Rivera, D. E., Atienza, A. A., Nilsen, W., Allison, S. M., & Mermelstein, R. (2011). Health behavior models in the age of mobile interventions: Are our theories up to the task? *Translational Behavioral Medicine*, 1(1), 53–71.
- Rothman, A. J., & Salovey, P. (1997). Shaping perceptions to motivate healthy behavior: The role of message framing. *Psychological Bulletin*, 121(1), 3–19.
- Schoeppe, S., Alley, S., Van Lippevelde, W., Bray, N. A., Williams, S. L., Duncan, M. J., & Vandelanotte, C. (2016). Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: A systematic review. *International Journal of Behavioral Nutrition and Physical Activity*, 13, 127.
- Scott, A. J., Webb, T. L., Martyn-St James, M., Rowse, G., & Weich, S. (2021). Improving sleep quality leads to better mental health: A meta-analysis of randomised controlled trials. *PLOS ONE*, 16(5), e0251956.
- Szinay, D., Jones, A., Chadborn, T., Brown, J., & Naughton, F. (2020). Influences on the uptake of and engagement with health and well-being smartphone apps: Systematic review. *Journal of Medical Internet Research*, 22(5), e17572.
- Torous, J., Andersson, G., Bertagnoli, A., Christensen, H., Cuijpers, P., Firth, J., ... & Areal, P. A. (2018). Towards a consensus on digital phenotyping. *World Psychiatry*, 17(1), 111–112.
- Torous, J., & Haim, A. (2018). Dichotomies in the development and implementation of digital mental health tools. *Psychiatric Services*, 69(12), 1204–1206.
- Villinger, K., Wahl, D. R., Boeing, H., Schupp, H. T., & Renner, B. (2019). The effectiveness of app-based mobile interventions on nutrition behaviours and nutrition-related health outcomes: A systematic review and meta-analysis. *NPJ Digital Medicine*, 2, 1–14.
- Fukuoka, Y., Gay, C. L., Joiner, K. L., & Vittinghoff, E. (2021). A novel diabetes prevention intervention using a mobile app: A randomized controlled trial with 12-month follow-up. *American Journal of Preventive Medicine*, 52(2), 223–231.

Torous, J., Lipschitz, J., Ng, M., & Firth, J. (2020). Dropout rates in clinical trials of smartphone apps for depressive symptoms: A systematic review and meta-analysis. *Journal of Affective Disorders*, 263, 413–419.