3-2-2022 2:15 PM

Examining the impact of financial incentive removal on physical activity: A quasi-experimental study of 584,760 mobile health application users

Sean Kevin Spilsbury, *The University of Western Ontario*

Supervisor: Mitchell, Marc S., *The University of Western Ontario*
Co-Supervisor: Prapavessis, Harry., *The University of Western Ontario*
A thesis submitted in partial fulfillment of the requirements for the Master of Arts degree in Kinesiology

Abstract

BACKGROUND: Government interest in using financial incentives (FIs) to stimulate physical activity (PA) is increasing. The cost of longer-term incentive interventions may be prohibitive, however. **PURPOSE:** To examine the impact of FI withdrawal on PA. **METHODS:** A 25-week retrospective pre-post quasi-experimental study was conducted with users of a FI-based mHealth app. Users from three Canadian provinces were included. Daily FI were removed in Ontario (ON; intervention) but not British Columbia (BC) and Newfoundland and Labrador (NL; control). Simple linear regression models were used to examine weekly mean daily step count after FI withdrawal. **RESULTS:** The total sample included 584,760 users (Female: 63.5%; Age: 34.3 years). Following FI withdrawal, weekly mean daily step count decreased in all provinces with the largest decrease observed in ON (i.e., 198 and 274 fewer steps/day vs. BC and NL, respectively). **CONCLUSION:** These findings may be relevant for governments looking to deploy time-limited FI-based PA programs.

Keywords

Financial Health Incentives, Financial Health Incentive Removal, Mobile Health, Smartphone Applications, Physical Activity, Behavioural Economics, Present Bias, Self-Determination Theory, Intrinsic Motivation, Extrinsic Motivation, Application Engagement, Transtheoretical Model

Summary for Lay Audience

To address the global physical inactivity pandemic, there is an urgent need for governments and corporations to implement sustainable and scalable population-level physical activity interventions. Incentive-based interventions delivered through smartphone apps can increase physical activity at the population-level and be cost-effective. However, effective strategies to remove financial incentives that maintain increases in physical activity are urgently needed for governments and corporations who cannot afford to continuously finance incentive-based interventions. This was a 25-week study that examined the impact of removing financial incentives for physical activity among 584,760 users of *Carrot Rewards*, a popular Canadian mobile health application. Users were categorized into subgroups to explore whether specific user characteristics influenced the impact of financial incentive removal on physical activity. Financial incentives for physical activity were removed in Ontario on Study Week 13 but were provided for 25-weeks in British Columbia and Newfoundland and Labrador. Declines in physical activity were greatest in Ontario relative to British Columbia and Newfoundland and Labrador. Furthermore, Ontario users who interacted with *Carrot Rewards* at the highest frequency and were the most physically active experienced the greatest decrease in physical activity after financial incentive removal. Length of exposure to *Carrot Rewards*, age and gender did not appear to influence the effect of financial incentive removal on physical activity. Given our study's sample size and real-world design, these findings may be applicable to governments and corporations with ongoing or planned incentive-based physical activity interventions delivered through smartphone applications at a population-level.

Acknowledgements

I would like to recognize everyone who supported me during my application to graduate school and the completion of my Master's degree. I decided to pursue graduate education to learn how to conduct research that improves health by shaping evidence-based policy and practice. I'd like to acknowledge those who helped me achieve this milestone in the penultimate requirement of my Master's degree, the written thesis. First, Dr. Marc Mitchell turned my hopes of an ideal graduate school experience into reality and has been an academic, professional, and personal mentor to me for over two years. I selected Dr. Mitchell to supervise my thesis because of his balance between research to advance knowledge and research to solve practical, real-world problems. However, I also believed he was professionally and personally relatable and would be a supervisor that could impart guidance for my career and life outside work. As his student for over two years, I can confidently say that Dr. Mitchell's teachings extend beyond academia and he has taught me to be a better pupil, professional, and person. I would hope that every graduate student could have the same supervisor-student relationship that I experienced with Dr. Mitchell. Thank you, Marc.

Second, I'd like to thank Dr. Harry Prapavessis and Dr. Piotr Wilk. As my co-supervisor, teacher, and manager, Dr. Prapavessis has improved my ability to conduct methodologically sound research, present information using oral and written communication, and teach students how to succeed at the undergraduate level. As a member of my thesis committee, Dr. Piotr Wilk helped me increase methodological rigor by providing guidance on study design, data analyses, and clarity of writing. Dr. Wilk

exceeded my expectations as a committee member by attending meetings outside of the committee and assisting with manuscript revision.

Third, I am thankful for the statistical assistance provided to me by Carolyn Taylor. In addition to her guidance on appropriate statistical analyses, Carolyn also expanded my knowledge of R, a programming language used for statistical computing. My experience in the use of R to organize and analyze the immense amount of data involved in this study will be invaluable to my career in professional research.

Fourth, I'd like to thank my academic colleagues who have helped both academically and personally in graduate school. Dave, I learned a great deal in first year watching you and Marc discuss your thesis on *Carrot Rewards* which taught me to be specific and concise when presenting the results of an incredibly large and complex study. Nabil, it was great to get the chance to work with you as a Teaching Assistant. Given the similarity of our research interests and personal interactions outside our graduate student responsibilities I hope to maintain contact with you going forward to see how we can collaborate professionally. Alex, you're a great friend and thoroughly hilarious. I'm glad we got the chance to meet one another and learn about shared interests. Our friendship could involve academia, but it doesn't which is why it's always refreshing to talk with you. Yoah, thank you for being my unofficial supervisor. I am grateful for your insight and assistance with my research, the wisdom you shared from your own experiences, and our philosophical discussions about subjects completely unrelated to our specializations. Madison, we started our Master's degree at the same time and shared memorable experiences that ranged from a field-trip to corporate offices in Toronto, completing a study within a week of meeting one another, and being classmates during first semester. However, I'm most grateful for the helping hand you and your Dad lent me while I was experiencing personal hardship. Your family helped me in the same way my own family would, and I am happy to say it changed my life. Thank you, Scott. Lastly, I'd like to thank Veronica for being such a great teammate who showed me how enjoyable collaborative research can be. I'm grateful I got the chance to co-author your publication; it was always fun to work with you and I appreciate the patience you exhibited when I couldn't dedicate time to our project while you could.

Fifth, I'd like to thank my friends and family who unconditionally supported my pursuit of higher education with enthusiasm and pride. Ben, you've been my best-friend since the first year of our undergraduate education when we were roommates in Saugeen. What stands out the most is your willingness to help me personally, academically, and professionally to this day. You gave me critical assistance with dataset organization for this study and facilitated the start of my education in R. By helping me utilize the tool to complete my thesis you also expanded my knowledge of programming languages which will be a fundamental and transferable skill in my professional career. Thanks, Ben. Tommy, I couldn't have asked for a better first-year graduate school roommate. I'm appreciative of your continued personal support in addition to what I've learned from you about professionalism, oral presentation, and abstract thinking. Thanks for taking the time to listen to my practice presentations and offer feedback to improve oral and visual delivery of information. To my Aunt Susan and Uncle Tony, thank you for assisting me in my graduate school application along with the feedback you've provided on the research I've conducted as a graduate student. To my brother Duncan, thank you for being the voice of reason who listened and provided objective opinions in response to my tendency to overthink. To my twin brother James, thank you for reminding me of my strengths and instilling me with confidence when I discussed the challenges of graduate school. To my sister Stephanie, thank you for reading my research and sharing your thoughts on where I could improve and incorporate new ideas. To my Dad, I can't express how grateful I am for your unwavering confidence in me to complete my Master's degree. However, I'm most appreciative of your flexibility to support my decision to attend university at the undergraduate and graduate level so I can pursue a meaningful career. To my Mom, I would not be where I am today without having you as a role model. You are the hardest worker I know which has taught me that effort is the best guarantee for success. I owe all my life accomplishments to work ethic which was instilled in me by you.

Finally, I'd like to thank Kassy Kristjanson for being by my side during the last and most challenging half of graduate school. Being able to share this personal milestone with you is more meaningful than any individual accomplishment of my life given everything we worked through to get to this point. I will be forever grateful for the chance you gave me to be a better person that has influenced where I am today. Thank you, Kassy.

Table of Contents

Abstract ii

Summary for Lay Audience iii

Acknowledgements iv

List of Tables xii

List of Figures xiii

List of Appendices xiv

Chapter 1 1

1 Introduction 1

1.1 Background 1

1.2 Mobile Health (mHealth) Interventions 2

1.3 Behavioural Economics and Financial Incentives 3

1.4 Incentive Removal and Sustained Physical Activity 5

1.5 Carrot Rewards 6

1.6 Study Purpose 7

Chapter 2 8

2 Literature Review 8

2.1 Physical Activity and Health 8

2.1.1 Recommendations for Health Benefits. 8

2.1.2 Good Practice Statements. 13

2.1.3 Interventions. 15

2.1.4 Evidence Gaps. 17

2.1.5 Recommendation Needs. 18

2.2 Digital Health Interventions and Physical Activity 21

2.2.1 mHealth. 21

2.2.2 Smartphone-based mHealth Interventions. 23

2.2.3 Attrition. 25

2.2.4 Engagement. 28

2.3 Incentive-based Interventions and Physical Activity 31

2.3.1 Incentive-based Digital Health Interventions and Engagement. 31

2.3.2 Incentive Design in Smartphone-Based mHealth Interventions. 36

2.4 Incentive Removal and Sustained Physical Activity 41

2.4.1 Contemporary Evidence. 41

2.4.2 Relevant Behaviour Change Theories. 43

2.4.3 Contradictions of ‘Undermined’ or ‘Crowded-out’ Intrinsic Motivation. 47

Chapter 3 52

3 Methodology 52

3.1 Setting 52

3.2 Program Description 52

3.3 Study Design 54

3.4 Outcome Measure 59

3.5 Covariates 60

3.6 Statistical Analyses 61

3.6.1 Sample Characteristics. 61

3.6.2 Primary and Secondary Analyses. 62

3.6.3 Complete Case Analysis. 65

3.6.4 Total Sample (Sensitivity) Analysis 67

Chapter 4 69

4 Results 69

4.1 Sample Characteristics 69

4.2 Primary Analyses 72

4.3 Secondary Analyses 76

4.4 Sensitivity Analyses 82

Chapter 5 85

5 Discussion 85

5.1 Main Findings 85

5.2 Practical Implications 86

5.3 Theoretical Implications 87

5.4 Comparison to Existing Literature 88

5.5 Limitations 92

5.6 Future directions 95

5.7 Conclusion 96

References 97

Appendices 131

Curriculum Vitae 165

List of Tables

Table 1. Baseline characteristics, complete cases sample. 70

Table 2. Pre-intervention period behaviours, complete cases sample. 71

Table 3. Estimated weekly mean daily step count intercepts and slopes (within provinces), complete cases sample. 74

Table 4. Estimated weekly mean daily step count intercepts and slopes (between provinces), complete cases sample. 75

Table 5. Estimated weekly mean daily step count intercepts and slopes by application engagement (within provinces), complete cases sample. 78

Table 6. Estimated weekly mean daily step count intercepts and slopes by application engagement (between provinces), complete cases sample. 80

Table 7. Estimated weekly mean daily step count intercepts and slopes (within provinces), total sample. 83

Table 8. Estimated weekly mean daily step count intercepts and slopes (between provinces), total sample. 84

List of Figures

Figure 1. Carrot Rewards “Steps” walking program interface. 54

Figure 2. Study flow chart. 58

Figure 3. Provincial weekly mean daily step count by week and intervention period, complete cases sample... 73

List of Appendices

Appendix A: Ethical Approval 131

Appendix B: Equations for the Primary and Secondary Analyses 132

Appendix C: Baseline Characteristics and Pre-Intervention Behaviours of the Total Study Sample and Users Excluded from the Complete Cases Sample 135

Appendix D: Application Engagement, Complete Cases Sample (Secondary Analyses) 141

Appendix E: Level of Physical Activity, Complete Cases Sample (Secondary Analyses) 142

Appendix F: Application Experience, Complete Cases Sample (Secondary Analyses) 148

Appendix G: Age, Complete Cases Sample (Secondary Analyses) 153

Appendix H: Gender, Complete Cases Sample (Secondary Analyses) 160

Appendix I: Curriculum Vitae 165

Chapter 1

1 **Introduction**

1.1 ***Background***

Physical inactivity (i.e., failure to meet physical activity [PA] guidelines; World Health Organization, 2020b) is a global pandemic (H. W. Kohl et al., 2012). Despite the irrefutable health benefits of regular physical activity PA (Warburton & Bredin, 2017), only half of Canadian adults are meeting the most recent PA guidelines of at least 150 minutes per week (min/wk) of moderate-vigorous intensity physical activity (MVPA; Ross et al., 2020; Statistics Canada, 2021a). Globally, data from population-based surveys indicate that 28% of adults fail to meet similar World Health Organization (WHO) guidelines (Guthold et al., 2018). Fortunately, recent evidence suggests that the health benefits of a physically active lifestyle are not only reserved for harder-to-achieve, higher-intensity MVPA (Chastin et al., 2019; Ekelund et al., 2019). Many health benefits are observed with regular light intensity PA (LPA; e.g., slower walking pace or light household chores) as well including reduced risk of depression (Mammen & Faulkner, 2013) and improved glycemic control (Chastin et al., 2019). Increases in time spent engaging in LPA could yield significant benefits for publicly-funded healthcare systems as well. For example, in Canada, a 1% reduction in the proportion of adults classified as “physically inactive” (i.e., < 5000 steps/day [steps/d]; Tudor-Locke et al., 2013) could generate $2.1 billion per year (Canadian) in direct healthcare system savings (Krueger et al., 2014). Effective and scalable interventions that increase population-level PA of any intensity, therefore, are needed (Füzéki et al., 2017; Reis et al., 2016).

1.2 *Mobile Health (mHealth) Interventions*

Digital or electronic health (eHealth) interventions involve the use of information and communications technologies to improve health and healthcare (World Health Organization, 2016). Mobile health (mHealth) interventions are a sub-segment of eHealth interventions that involve the use of mobile devices (i.e., smartphones and wearable trackers) to improve health and healthcare (World Health Organization, 2016). Contemporary smartphones include ‘built-in’ accelerometers that capture PA data which can be used by mHealth applications (apps) to deliver more personalized PA interventions (Harari et al., 2016). The scalability potential of mHealth-based PA interventions is high with smartphone ownership, for instance, approaching 90% in Canada and the U.S. (Pew Research, 2021; Statistics Canada, 2021b). Mobile health apps that promote PA are becoming increasingly popular with more than 100,000 in the major app stores (Research2Guidance, 2017). Furthermore, PA app supply (approximately 8,000 more PA apps published in 2020 vs. 2019; Sydow, 2021) and demand (approximately 600 million more PA app downloads in 2020 vs. 2019; Sydow, 2021) has grown considerably of late, especially since the beginning of the COVID-19 pandemic (Government of Canada, n.d.-a) in part reflecting physical distancing policies (Government of Canada, n.d.-b; Newbold et al., 2021).

Meta-analytic evidence from randomized controlled trials (RCTs) suggests that mHealth interventions can increase PA (i.e., 1566 to 1850 steps/d in interventions up to two years long; Laranjo et al., 2021; Mönninghoff et al., 2021). Metaregression models suggest that mHealth intervention features that increase user engagement (i.e., the amount, frequency, duration, and depth of usage; Perski et al., 2017) and retention (i.e., percentage returning for follow-up assessment; Laranjo et al., 2021) are consistently associated with greater intervention effectiveness (Laranjo et al., 2021). Examples of such features include individualized goal setting, timely biofeedback, and opportunities to connect with similar others (Mitchell, Orstad, et al., 2020). Unfortunately, a recent examination of 1000 PA apps suggests that retention rates (measured by the percentage of users that return to an app after their last visit) over 90 day and one year periods are only 31% and 19%, respectively (Apptentive, 2021), limiting their behaviour change potential. Furthermore, PA improvements often wane several weeks and months after interventions are discontinued (i.e., by about 700 steps/d; Mönninghoff et al., 2021). In addition to considering promising mHealth intervention features, behavioural economics (BE), the Nobel Prize winning (2017; The Nobel Prize, 2017) offshoot of traditional economic theory that incorporates insights from psychology, has emerged as a theoretical framework from which other practical solutions to the mHealth app engagement/retention issue could arise (Mitchell, Orstad, et al., 2020).

1.3 *Behavioural Economics and Financial Incentives*

Behavioural economics describes how systematic errors in human decision making, called “decision biases”, can lead to poor health-related decisions and adverse health outcomes (Thaler & Sunstein, 2008). The “present bias”, for example, describes the human tendency to overemphasize the current ‘costs’ (e.g., time out of a busy schedule) of a health behaviour relative to the future ‘benefits’ of that behaviour (e.g. improved health and longevity; Camerer & Loewenstein, 2003). By leveraging an individual’s tendency to act in favour of their immediate self-interest, BE postulates that the immediate provision of a financial incentive (FI) for engaging in PA may encourage more people to participate today, rather than put it off until tomorrow (M. A. Adams et al., 2017; Loewenstein et al., 2013). Financial incentives for health are rewards with monetary value that are contingent on the achievement of a pre-specified health behaviour or outcome (J. Adams et al., 2014) such as walking more (Strohacker et al., 2014) or losing weight (Burns et al., 2012). Indeed, government (i.e., Health Incentives Scheme, United Kingdom; Department of Health and Social Care & Churchill, 2021), corporate (i.e., 52% of employers in 2019; Willis Towers Watson, 2020), and public-private (StepUp Program with the University of Pennsylvania and 24 Hour Fitness; Milkman et al., 2021) interest in FIs for health is rising.

Evidence from recent meta-analyses suggests that FI provision may improve PA in short- (< 6 months) and long-term (> 6 months) interventions (607 - 754 steps/d; Luong et al., 2020; Mitchell, Orstad, et al., 2020). On a population-level though, the costs associated with ongoing FI provision can be prohibitive for governments and corporations looking to deploy this kind of intervention as efficiently as possible (Rondina et al., 2021). Interestingly, a systematic review by Mitchell, Orstad, et al. (2020) suggests that more expensive, indefinite FI provision may not be necessary. For example, their pooled analyses suggest that PA improvements may persist three to six months after FI removal (i.e., 514 steps/d; Mitchell, Orstad, et al., 2020). On the contrary though, Mitchell, Orstrad, et al. (2020) indicate in their narrative evidence summary (using ‘vote counting’) that only 22% (n = 4/18) of included RCTs reported post-intervention PA increases (Mitchell, Orstad, et al., 2020). These inconsistent findings suggest more research is needed to uncover the impact of FI removal on PA in different, fiscally constrained public health settings.

1.4 *Incentive Removal and Sustained Physical Activity*

According to self-determination theory (SDT), motivation is reflected along an internalization continuum that represents the degree to which behaviour is self-determined (i.e., internalized; R. M. Ryan & Deci, 2000). Amotivation (i.e., lack of motivation) and intrinsic regulation (i.e., intrinsic motivation) lie at each end of the continuum and are separated by four types of extrinsic motivation (i.e., external, introjected, identified, and integrated regulation; R. M. Ryan & Deci, 2000). Behaviour becomes more self-determined when one moves from amotivation to intrinsic regulation along the continuum and position is determined by the extent to which basic psychological needs for competence, autonomy and relatedness are satisfied by social contexts (R. M. Ryan & Deci, 2000). Findings from years of psychology research suggests that intrinsic motivation measured by time spent on enjoyable tasks (e.g., completing puzzles) declines in response to extrinsic reward provision (Deci et al., 1999). Decreased intrinsic motivation after the receipt of extrinsic rewards has been defined as the “undermining effect” in traditional SDT-grounded psychology research (Deci et al., 1999) and “crowding out” in economic literature (Frey & Jegen, 2001). It has also been suggested, however, that deploying multiple theories of behaviour change in the development of mHealth interventions may in fact protect intrinsic motivation (J. M. Murray et al., 2020). A recent mediation analysis of a digital PA intervention informed by learning theory, social-cognitive theory and SDT, for example, suggested that assignment to receive FIs for PA goal achievements increased intrinsic motivation relative to non-incentive controls. In addition, increases in more self-determined forms of motivation (i.e., integrated regulation, intrinsic motivation) were associated with improved PA at intervention end (Study Month 6) and six-month follow-up (Study Month 12; J. M. Murray et al., 2020). Other theoretically-grounded RCT studies report similar results whereby FI provision, in combination with other behaviour change techniques (e.g., goal setting, action planning, self-monitoring, etc.), led to maintained or increased self-determined motivation as well as improved PA post-intervention (Budworth et al., 2019; Kramer et al., 2020; J. M. Murray et al., 2019). Specifically, it has been postulated that extrinsic rewards (e.g., FIs) may actually increase intrinsic motivations if they engender feelings of perceived competence, one of SDTs three basic psychological needs (R. M. Ryan & Deci, 2000). Given important practical implications of choosing to withdraw FIs (or not) after a period of time in the context of population-level PA programming, more research is needed that examines PA behaviours after FI removal in real-world settings.

1.5 ***Carrot Rewards***

Scientific advances in the mHealth field through systematic exploration of commercially available PA app data may improve population-level PA (2018 Physical Activity Guidelines Advisory Committee, 2018). *Carrot Rewards* was a free mHealth app that rewarded Canadians for engaging in healthy behaviours such as walking (Public Health Agency of Canada, 2015). Grounded in BE (Camerer & Loewenstein, 2003) and SDT (Deci et al., 1999), the multi-component *Carrot Rewards* app was downloaded by over 1.3 million Canadians and reported more than 500,000 monthly active users as of May 2019 (Pearson et al., 2020). The app provided FIs in the form of loyalty reward points (redeemable for consumer goods like groceries or gas) for completing PA goals (i.e., daily, weekly and team-based goals; Mitchell et al., 2017), and was available in the provinces of Ontario (ON), British Columbia (BC) and Newfoundland and Labrador (NL) between 2016 and 2019 (Marotta, 2019). On December 8, 2018, rewards for daily step goal completion were discontinued in ON due to a lack of funding, but not in BC or NL (Ng, 2018). The partial withdrawal of FIs in ON (i.e., daily step count rewards were removed while longer-term, harder-to-achieve team-based rewards persisted; Ng, 2018) provides a unique opportunity to investigate the impact of FI removal on PA in a real-world, quasi-experimental context.

1.6 ***Study Purpose***

The overarching purpose of this research, then, is to examine the effect of partial FI removal on PA in a population-level context. The primary study objective is to examine the impact of daily FI removal on weekly mean daily step counts in ON compared to BC and NL where FI availability did not change. Secondary objectives are to explore whether co-variates (e.g., PA and mHealth app engagement levels, age) influence the impact of daily FI removal on weekly mean daily step counts in ON compared to BC and NL.

Chapter 2

2 Literature Review

2.1 *Physical Activity and Health*

2.1.1 ***Recommendations for Health Benefits.*** The 2020 WHO Guidelines on Physical Activity and Sedentary Behavior (GPASB) provide the latest evidence-based public health recommendations for children, adolescents, adults, and older adults with or without a chronic condition and/or disability, and pregnant and postpartum women on the amount of PA (frequency, intensity, duration) necessary for significant health benefits and reduced health risks. The 2020 WHO GPASB expanded on the 2010 WHO Global Recommendations on Physical Activity for Health (GRPAH) by utilizing and systematically updating evidence collected in the development of recent national PA guidelines such as Canada, Australia and the United States (World Health Organization, 2010). Inclusion criteria required that reviews be conducted in accordance to standard systematic processes with sufficient literary documentation, assessed for certainty of evidence using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) procedure or an equivalent methodology, and address populations of interest with no restrictions to country or income level (Bull et al., 2020). For the guidelines on children, adolescents, pregnant women and all other age and subpopulation groups, systematic reviews and a scientific report that informed national public health policy for PA were used and updated (2018 Physical Activity Guidelines Advisory Committee, 2018; Carson et al., 2016, 2017; Mottola et al., 2018). To update evidence, a search for systematic reviews and pooled analyses of cohort studies was conducted for research published from the last search date in the included national PA guidelines to September 2019 (World Health Organisation, 2020b).

The Guideline Development Group (GDG) considered wording and evidence strength rating to formulate recommendations based on the balance of benefits to harms, the certainty of evidence, value and preference sensitivity, the potential impact on gender, social and health equity, and acceptability, feasibility and resource implications (Bull et al., 2020). Population, intervention/exposure, comparison, and outcome (PI/ECO) questions for each subpopulation addressed the association between PA and health-related outcomes, if there was a dose-response (volume, duration, frequency, intensity) relationship, and whether associations varied by type or domain (leisure time, occupational, transportation, household, and education) of PA. The GRADE procedure was implemented to rate the certainty of evidence for each PI/ECO question and yielded quality ratings of very low, low, moderate or high (Balshem et al., 2011; Guyatt et al., 2008). The GDG evaluated the totality of evidence for each recommendation and assigned grades of strong if the balance of benefits to harms was assessed as substantial for the target population and conditional if the balance of benefits to harms was small or there was significant variability in benefits to the target population. The 2020 WHO GPASB are applicable to all populations across the age groups of 5 years and above, regardless of gender, cultural background or socioeconomic status and are relevant to individuals of all abilities. Those with chronic conditions and/or disability along with pregnant and postpartum women should strive to meet recommendations when possible and if capability permits (World Health Organisation, 2020b). The following PA recommendations associated with health-related outcomes are provided for population subgroups that are substantiated by recommendation strength and certainty of evidence along with good practice statements to guide implementation. The 2019 WHO Guidelines on Physical Activity, Sedentary Behavior, and Sleep for Children Under 5 Years of Age (GPASBSC) not included in the 2020 WHO GPASB are also outlined (World Health Organization, 2019).

The 2019 WHO GPASBSC Under 5 Years of Age indicates that health benefits of PA can begin from birth. Though overall quality of evidence is very low, there is moderate quality evidence for cognitive development, low quality evidence for psychosocial health, motor development and adiposity and very low-quality evidence for fitness. Infants (less than 1 year) should perform PA several times a day in a variety of forms, particularly through floor-based play and more is better. Infants who are not mobile should spend at least 30 minutes a day in prone position (i.e., tummy time) while awake. Children 1-2 years of age should engage in at least 180 minutes of any intensity PA that varies in format including MVPA during the day. Children 3-4 years of age should spend at least 180 minutes per day (min/d) in a variety of PA at any intensity, of which at least 60 minutes is MVPA (strong recommendation, very low quality evidence; World Health Organization, 2019). According to the 2020 WHO GPASB, health benefits from PA in children and adolescents (aged 5-17 years) include improved physical fitness (cardiorespiratory and muscular fitness), cardiometabolic health (blood pressure, dyslipidemia, glucose, and insulin resistance), bone health, cognitive outcomes (academic performance and executive function), mental health (reduced symptoms of depression), and decreased adiposity. For children and adolescents living with a disability, PA can improve cognition among individuals with diseases or disorders that impair cognitive function including attention-deficit/hyperactivity disorder and may increase physical functioning in children with an intellectual disability. Children and adolescents with or without one of the stated disabilities should undertake an average of at least 60 min/d of primarily aerobic MVPA throughout the week. Aerobic vigorous-intensity (≥ 6 METs) PA (VPA) as well as muscle and bone strengthening activities should be performed at least 3 days a week (strong recommendation, moderate certainty evidence; World Health Organization, 2020b).

In adults (aged 18-64 years) and older adults (aged 65 years and over), PA leads to health benefits such as improved all-cause mortality, cardiovascular disease mortality, incident hypertension, incident site-specific cancers (bladder, breast, colon, endometrial, oesophageal adenocarcinoma, gastric and renal), incident type 2 diabetes, mental health (reduced symptoms of anxiety and depression), cognitive health, sleep, and adiposity. Physical activity also helps prevent falls, fall-related injuries and declines in bone health and functional ability among older adults. All adults and older adults should participate in regular PA and engage in at least 150-300 minutes of aerobic moderate-intensity (3 > 6 METs) PA (MPA), 75-150 minutes of aerobic VPA, or an equivalent combination of both intensities per week for substantial health benefits (strong recommendation, moderate certainty evidence). Additional health benefits are accrued by performing muscle-strengthening activities of at least moderate-intensity that involve all major muscle groups on 2 or more days a week (strong recommendation, moderate certainty evidence). To prevent falls and enhance functional capacity, older adults should do varied multicomponent PA that focuses on balance and strength training of at least moderate intensity on 3 or more days a week (strong recommendation, moderate certainty evidence). Adults and older adults may increase aerobic MPA to more than 300 minutes, engage in more than 150 minutes of aerobic VPA or an equivalent combination of both intensities throughout the week for further health benefits (conditional recommendation, moderate certainty evidence; World Health Organization, 2020b).

Physical activity also provides health benefits to adults and older adults living with chronic conditions and/or disabilities. In cancer survivors, PA improves all-cause mortality, cancer-specific mortality, and risk of cancer recurrence or second primary cancer. For individuals living with hypertension, PA improves cardiovascular disease mortality, disease progression, physical functioning, and health-related quality of life. PA reduces rates of mortality from cardiovascular disease and indicators of disease progression in people living with type 2 diabetes mellitus. Among individuals living with HIV, PA can improve physical fitness and mental health (reduced symptoms of anxiety and depression) and does not negatively affect disease progression (CD4 count and viral load) or body composition. In adults living with multiple sclerosis, PA can improve physical function and physical, mental, and social domains of health-related quality of life. For individuals with a spinal cord injury, PA can improve walking function, muscular strength, upper extremity function, and enhance health-related quality of life. In individuals with diseases or disorders that impair cognitive function, PA can improve physical functioning and cognition in people with Parkinson's disease and a history of stroke, may increase quality of life among adults with schizophrenia and enhance physical function for adults with intellectual disability, and augment quality of life in adults with major clinical depression. When not contraindicated by the stated chronic conditions and/or disabilities, all adults and older adults with the listed chronic conditions and/or disabilities should adhere to PA recommendations for adults and older adults. Health benefits of equal magnitude are achieved and supported by the same recommendation strengths and evidence certainty ratings (World Health Organisation, 2020b).

In pregnant and postpartum women, PA during pregnancy and the postpartum period leads to maternal and fetal health benefits such as decreased risk of pre-eclampsia, gestational hypertension, gestational diabetes, excessive weight gain, delivery complications, postpartum depression, fewer newborn complications, no adverse effect on birthweight, and no increase in risk of stillbirth. Pregnant and postpartum women without contraindications should participate in regular PA throughout pregnancy and the postpartum period, perform at least 150 minutes of aerobic MPA during the week for substantial health benefits, and incorporate a range of aerobic and muscle-strengthening activities including gentle stretching (strong recommendation, moderate certainty evidence). In addition, women who routinely engaged in aerobic VPA or were regularly physically active prior to pregnancy can continue these activities during pregnancy and the postpartum period (strong recommendation, moderate certainty evidence; World Health Organization, 2020b).

2.1.2 ***Good Practice Statements.*** The 2020 WHO GPASB also provides good practice statements that are not graded recommendations but are derived from scientific evidence and practical considerations reviewed and endorsed by the GDG. In general, good practice statements are similar for children, adolescents, adults, and older adults with or without a disability and/or chronic illness along with pregnant and postpartum women. Performing some PA is better than none, those not meeting recommendations can acquire health benefits by participating in any PA and should start by engaging in small amounts of PA that gradually increases in frequency, intensity, and duration over time (World Health Organisation, 2020b).

It is critical to provide children and adolescents with safe and equitable PA opportunities and encourage participation in physical activities that are enjoyable, offer variety, and are appropriate for their age and ability. Older adults should be as physically active as their functional ability permits and adjust PA effort levels relative to degree of fitness. Adults with chronic conditions can consult with a PA specialist of health-care professional for advice on the type and amount of activity suited for their individual needs, abilities, functional limitations/complications, medications, and overall treatment plan. Medical clearance is generally not necessary for individuals without contraindications prior to adopting LPA or MPA that doesn’t exceed the demands of brisk walking and activities of daily living. There are not major risks for children, adolescents, and adults living with a disability performing PA when it is appropriate to their current activity level, health status, and physical function given that the health benefits gained outweigh the risks. Children, adolescents, and adults living with a disability may require consultation with a health-care professional or a PA and disability specialist to help determine appropriate types and amounts of activity suitable for their individual needs (World Health Organisation, 2020b).

Pregnant women may perform pelvic muscle floor training daily to reduce risk of urinary incontinence. Additional safety considerations for pregnant women participating in PA include: avoid PA in excessive heat, particularly when humidity is high; maintain hydration before, during and after PA; avoid PA that involves physical contact, has a high risk of falling, or limits oxygenation; avoid PA in the supine position after the first trimester; seek supervision from a specialist health-care provider when considering athletic competition or exercise that significantly exceeds recommended guideline; be informed by their health-care provider of danger signs of when to stop, or limit PA and consult a qualified health-care provider immediately if they occur; return to PA gradually after delivery, and in consultation with a health-care provider in the case of delivery by Caesarean section (World Health Organisation, 2020b).

2.1.3 ***Interventions.*** Unlike the 2020 WHO GPASB, the 2018 Physical Activity Guidelines Advisory Committee (PAGAC) Scientific Report includes a review of evidence for interventions designed to supplement knowledge with specific approaches and strategies that effectively promote and sustain PA. Methodological quality of systematic reviews and meta-analyses was assessed using a modified version of A Measurement Tool to Assess Systematic Reviews (AMSTAR; Johnson et al., 2014; Shea et al., 2007). Risk of bias was assessed using an adapted version of the United States Department of Agriculture (USDA) Nutrition Evidence Library (NEL) Bias Assessment Tool (BAT; Office of Disease Prevention and Health Promotion, 2015). Evidence was graded using a rubric adapted from the USDA NEL Conclusion Statement Evaluation Criteria to reflect the specific characteristics of PA literature (Office of Disease Prevention and Health Promotion, 2015). Strength of evidence was graded as strong, moderate, limited or not assignable and based on applicability of the populations, exposures and outcomes studied, generalizability to the populations of interest, risk of bias and limitations, quantity and consistency of findings across studies, and the magnitude and precision of effect (2018 Physical Activity Guidelines Advisory Committee, 2018).

The 2018 PAGAC Scientific Report found that “efforts to promote physical activity can be effective” (A-5). Individual-level interventions can increase volumes of PA in youth and adults, particularly when interventions are informed by behaviour change theories and techniques (strong evidence). Multi-component school-based programs and community-wide interventions that extensively contact the majority of targeted populations are effective at improving levels of PA (strong and moderate evidence, respectively). Environmental and policy changes such as modifying built environments to induce PA (i.e., physically active transport) are positively associated with increased walking and cycling compared to areas that lack these features (moderate evidence). Wearable activity monitors when used in conjunction with behavioural change strategies such as goal setting can improve PA among general adult populations as well as those with type 2 diabetes (strong evidence). Telephone assisted interventions lasting at least one year can enhance PA among general adult populations and older adults (strong evidence). Internet-delivered interventions that include educational components can increase levels of PA in the general adult population (strong evidence). Computer-tailored print interventions that collect user information through mailed surveys to generate personalized advice and support have a small but positive effect on increasing levels of PA in general adult populations (moderate evidence). Mobile phone programs that consist of or include text-messaging have a small to moderate effect on enhancing levels of PA in general adult populations and use of smartphone applications can increase regular PA in children and adolescents (moderate and strong evidence, respectively; 2018 Physical Activity Guidelines Advisory Committee, 2018).

2.1.4 ***Evidence Gaps.*** Research needs identified by the 2020 WHO GPASB indicate a lack of information across population subgroups on more precise details of the dose-response relationship between PA and several health outcomes studied, the health benefits of LPA, the differences in health effects from different types and domains of PA and the joint association between PA and sedentary time with health outcomes across the life span. There is limited evidence from low- and middle-income countries, economically disadvantaged or underserved communities, and in people living with disability and/or chronic disease. Oftentimes, studies are not designed or powered to test for effect modification by sociodemographic information (age, sex, race/ethnicity, socioeconomic status) that may modify the health effects of PA. Research on PA and health outcomes that consider vulnerable populations and sociodemographic characteristics are important for increasing the specificity of public health recommendations and reducing health disparities (World Health Organisation, 2020b).

The 2018 PAGAC Scientific Report outlined similar evidence gaps but also provided research recommendations for PA interventions. Effective intervention strategies need to be identified to increase PA in multiple settings among diverse population subgroups. Determining how intervention effectiveness differs by sociodemographic characteristics is also critical. To develop effective population-level PA interventions that improve public health, diverse subgroups must be included in research designs. Data collected across population subgroups can inform formative design methods and increase intervention effectiveness by targeting needs of specific subgroups along with individual preferences and requirements (2018 Physical Activity Guidelines Advisory Committee, 2018). Interventions that received a strong or moderate evidence grade need to develop and systematically test methods to effectively implement PA promotion techniques in real-world settings. Given that 27.5% of adults and 81% of adolescents did not meet the 2010 WHO GRPAH recommendations in 2016 (Guthold et al., 2018, 2020; World Health Organization, 2010), development and systematic testing of potentially effective methods and techniques in population-level PA interventions is critical for public health (2018 Physical Activity Guidelines Advisory Committee, 2018). Lastly, further exploration of methods and pathways to systematically exploit the extensive amount of commercially available data and interventions relevant to PA is necessary. As of 2019, nearly a third of the global population own smartphone devices and built-in accelerometers can accurately track step count (iPhone and Android; Duncan et al., 2017; Hekler et al., 2015; Taylor & Silver, 2019). Averaging 7000 steps/d is consistent with obtaining at least 150 minutes of accelerometer measured MVPA per week and dose-response evidence indicates a linear relation of daily step count with all-cause mortality, cardiovascular disease and type 2 diabetes mellitus (Kraus et al., 2019; Tudor-Locke et al., 2011). Systematic understanding to appropriately use naturally-occurring PA databases may improve population-level intervention effectiveness and increase public health benefits (2018 Physical Activity Guidelines Advisory Committee, 2018).

2.1.5 ***Recommendation Needs.*** Although LPA is endorsed by the 2020 WHO GPASB (World Health Organisation, 2020b), no recommendations for the amounts necessary to obtain health benefits are provided despite moderate to high certainty evidence of reduced all-cause mortality (Amagasa et al., 2018; Ekelund et al., 2019; World Health

Organisation, 2020a). Light-intensity physical activity recommendations are important for PA promotion to inform individuals about attainable health benefits if they are unable to initially perform MVPA (i.e., contraindications, mobility limitations) or are unwilling to participate in PA at higher intensities due to feelings of discomfort (Qiu et al., 2021). Amagasa et al. (2018) conducted a systematic review and found that replacing 30-60 minutes of sedentary behaviour (SB; ≤ 1.5 MET) with LPA was associated with lower risk of all-cause mortality after adjustment for MVPA (Hazard ratio [HR] 0.80 – 0.88, 95% Confidence Interval [CI] 0.73 – 0.92) which was graded as moderate certainty evidence (Amagasa et. al., 2018; World Health Organisation, 2020a). In a meta-analysis by Ekelund et al. (2019), a dose-response relationship was demonstrated between LPA and all-cause mortality across four quartiles of increasing LPA. Each quartile corresponded to approximately 199.5 (referent quartile), 258.5, 308.5, and 379.5 min/d of LPA, respectively. Risk for all-cause mortality decreased across each quartile of increased LPA (Second quartile: HR 0.66, 95% CI 0.58 – 0.74; Third quartile: HR 0.51, 95% CI 0.44 – 0.57; Fourth quartile: HR 0.44, 95% CI 0.34 – 0.59) and was graded as high certainty evidence (Ekelund et al., 2019; World Health Organisation, 2020a).

Since the 2020 WHO GPASB, health benefits from LPA have been demonstrated in meta-analyses that studied effects independent from, and comparative to, MVPA (Ku et al., 2020; Qiu et al., 2020, 2021). Using meta-regression models, Ku et al. (2019), found a significant (p = .012) log-cubic dose-response relationship ($\beta = -0.78^{-3}$; standard error [SE], 0.31^{-3}) between objectively measured daily LPA and mortality in adults and older adults, independent of MVPA (Ku et al., 2019). Qiu et al. (2020) investigated the association of objectively measured LPA with risk of cancer mortality in the general population. Comparisons between the effectiveness of LPA and MVPA in reducing cancer mortality were conducted to promote use of LPA in clinical practice. Light-intensity physical activity for 30 min/d decreased risk of cancer mortality by 14% (pooled HR 0.86, 95% CI 0.79 – 0.95; $I^2 < 1\%$) and the dose-response analysis indicated this relationship was linear ($p_{non\text{-}linearity} = 0.72$). Comparable magnitudes in risk reduction of cancer mortality were demonstrated between LPA (HR 0.87, 95% CI 0.79 – 0.97) and MVPA (HR 0.94, 95% CI 0.79 – 1.13) for equal time length (30 min/d) that were not significantly different ($p_{\text{interaction}} = 0.46$). Magnitudes in risk reduction of cancer mortality were also similar between LPA (HR 0.74, 95% CI 0.59 – 0.93) and MVPA (HR 0.94, 95% CI 0.79 – 1.13) for equal activity amount (150 MET min/d) and not significantly different ($p_{\text{interaction}} = 0.11$; Qiu et al., 2020). Qiu et al. (2021) also examined the association of objectively measured LPA with risk of cardiovascular mortality in the general population. Similar to Qiu et al. (2020), the effectiveness of LPA and MVPA to reduce cardiovascular mortality was compared to facilitate use of LPA in clinical practice. Daily LPA for 30 minutes reduced risk of cardiovascular mortality by 20% (pooled HR 0.80, 95% CI 0.67–0.96) although evidence of heterogeneity was significant ($I^2 = 84\%$) and the dose-response analysis suggested a non-linear relationship ($p_{\text{non-linearity}} = 0.004$). Comparison by equal activity amount (150 MET min/d) was not significantly different ($p_{\text{comparison}} = 0.41$) between LPA (65 min/d) and MVPA (30 min/d) in reducing cardiovascular mortality (HR 0.67, 95% CI 0.48 – 0.93 and HR 0.54, 95% CI 0.37 – 0.81, respectively; Qiu et al., 2021). Given the health benefits from regular LPA (e.g., all-cause mortality) and barriers associated with MVPA (e.g., contraindications), it is clinically important to recommend daily LPA. Interventions designed to increase LPA such as step counting devices and mHealth apps should be promoted at the population-level (Qiu et al., 2021).

2.2 *Digital Health Interventions and Physical Activity*

2.2.1 *mHealth.* Digital health interventions (DHI) describe electronic health (eHealth) and mHealth treatments, where the former involves the use of mobile technologies such as phones, tablets, and tracking devices to aid and improve public health practice (World Health Organization, 2016). The WHO identified mHealth as an important part of a comprehensive “systems-based” solution to achieve global physical inactivity targets (15% reduction by 2030) in their Global Action Plan on Physical Activity 2018 – 2030 (World Health Organization, 2018). In a recent meta-analysis, Mönninghoff et al. (2021) examined the immediate, short-term, and long-term effectiveness of mHealth interventions on PA. Investigations as to whether effects differed by population subgroup (healthy, at-risk, or sick), intervention design (scalable; no human-to-human interaction versus nonscalable; human-to-human interaction) and type of control group (nonmobile, information material only, or no intervention) were also conducted. Eligible outcomes were walking, MVPA, total physical activity (TPA) and energy expenditure (EE). At end of intervention, significant increases were demonstrated for walking (standard mean difference [SMD] 0.46, 95% CI 0.36 - 0.55; p <.001), MVPA (SMD 0.28, 95% CI 0.21 - 0.35; p <.001), TPA (SMD 0.34, 95% CI 0.20 - 0.47; p <.001), and EE (SMD 0.44, 95% CI 0.13 - 0.75; p = 0.005). Short-term effects were sustained (≤ 6 months after end of intervention) for walking (SMD 0.26, 95% CI 0.09 - 0.42; p = 0.002), MVPA (SMD 0.20, 95% CI 0.05-0.35; p = 0.008), and TPA (SMD 0.53, 95% CI 0.13 - 0.93; p = 0.009). Long-term (> 6 months after end of intervention) were sustained for walking

(SMD 0.25, 95% CI 0.10 - 0.39; p = 0.001) and MVPA (SMD 0.19, 95% CI 0.11 - 0.27; p <.001). Study population was an effect moderator, with higher effect scores in sick and at-risk subgroups compared to healthy populations. Scalable and non-scalable mHealth interventions significantly increased PA at similar levels. Mobile health interventions led to increased walking, MVPA, and TPA when compared to studies using nonmobile treatments, information material only or no intervention (Mönninghoff et al., 2021).

Mönninghoff et al. (2021) conducted one of the first analyses to indicate that mHealth PA interventions are superior to nonmobile treatments. However, findings must be interpreted cautiously given the high risk of bias in 80.3% (94/117) of included studies and significant heterogeneity resulting in very low to low quality evidence (Balshem et al., 2011; Sterne et al., 2019). Long-term effectiveness evidence was comprised of follow-up measurements taken on average 13.96 months post-intervention in only 8 studies and effect sizes diminished from almost moderate at end of intervention to small at longest follow-up (Mönninghoff et al., 2021). Furthermore, the effectiveness of delivery methods was not examined as pedometers or accelerometers with displays, activity trackers, smartphones, and tablets, were included in the definition of mHealth interventions. Laranjo et al. (2021) conducted a systematic review and meta-analysis to examine the effectiveness of PA interventions using smartphone apps or activity trackers with automated and continuous self-monitoring and feedback in adults (aged 18 – 65 years) without chronic disease. Results demonstrated that interventions using smartphone apps or activity trackers had a positive effect on PA at a mean follow-up of 13-weeks compared with control conditions (SMD 0.350, 95% CI 0.236 – 0.465; $p < 0.0001$, I^2 = 69%, T^2 = 0.051) corresponding to an increase of 1850 steps/d. Significant effects were found in subgroup analyses of interventions using goals and planning (SMD 0.446, 95% CI 0.33 - 0.562, $p < 0.0001$), graded tasks (SMD 0.512, 95% CI 0.337 - 0.687, $p = 0.031$), text messaging (SMD 0.495, 95% CI 0.335 - 0.654, $p = 0.028$), personalization (SMD 0.541, 95% CI 0.365 - 0.718, $p = 0.006$), and behaviour change theories (SMD 0.449, 95% CI 0.312 - 0.587, $p = 0.018$). Subgroup metaregression indicated that text messaging, personalization, and retention rate were significantly associated with intervention effectiveness (i.e., accounted for 71% of the variance in intervention effectiveness; $R^2 = 0.71$). Notably, study duration was not associated with intervention effectiveness. In addition, there were no significant differences in intervention effectiveness between studies that used smartphone apps or activity trackers (Laranjo et al., 2021). Lastly, Mitchell, Orstad, et al. (2020) conducted a meta-analysis and systematic review to examine the short- (< 6 months) and long-term (≥ 6 months) effects of FIs on daily step count. Secondary objectives were to determine whether PA persisted after FI removal and reduce heterogeneity between studies through subgroup meta-analyses. In contrast to findings from Laranjo et al. (2021), subgroup analyses indicated that studies which used wearable activity trackers outperformed studies that employed smartphones during intervention (834 steps/d) and post-intervention follow-up (620 steps/d; Mitchell, Orstad, et al., 2020).

2.2.2 ***Smartphone-based mHealth Interventions.*** Mobile health interventions delivered through smartphones can utilize mobile sensor data to accurately measure step count with built-in accelerometers (iPhone and Android; Duncan et al., 2017; Hekler et al., 2015). Step counts can be translated to standard PA guidelines (Tudor-Locke et al., 2013) and smartphones can schedule delivery of intervention content to account for time of day and momentary environments of users (Harari et al., 2016). Furthermore, smartphone technologies can provide high-level personalization for users that collect behavioural data unobtrusively and on site (Harari et al., 2016). Thus, smartphone-based mHealth interventions for PA are accessible, scalable, and comparatively inexpensive to treatments requiring human-to-human interaction (Domin et al., 2021). A limited number of systematic reviews and meta-analysis specific to smartphone-based PA mHealth interventions have demonstrated mixed evidence of effectiveness (Bort-Roig et al., 2014; Feter et al., 2019; Romeo et al., 2019). Bort-Roig et al. (2014) conducted one of the first systematic reviews on the use of smartphones in PA measurement and promotion. The aim of the study was to examine the extent to which smartphones could be effectively used to measure and influence PA. Findings from 17 studies which implemented and evaluated a smartphone intervention indicated that PA profiles, goal setting, real-time feedback and online expert consultation were the most useful techniques to encourage PA change. Of the five studies that assessed intervention effectiveness, four reported increased PA (800 – 1104 steps/d) ranging from two weeks to six months and one demonstrated maintenance (> 10,000 steps/d) over three months (Bort-Roig et al., 2014). More recently, Feter et al. (2019) performed a meta-analysis to examine the effectiveness of smartphone-based interventions in PA promotion. Randomized and non-randomized studies with PA interventions that used either text-messaging or an app to promote PA in adults were included in the meta-analysis. Results indicated that smartphone-based mHealth interventions led to increased PA by 12.02 min/d (95% CI 5.45 – 18.60; $p < .001$) and 1999.59 steps/d (95% CI 1036.49 – 2962.69; $p < 0.001$) compared to control conditions without a smartphone. App-specific smartphone-based mHealth interventions had a significant positive effect on the number of steps (SMD 0.18, 95% CI 0.01 – 0.35; *p* = 0.04) and min/d of PA (SMD 0.31, 95% CI 0.01 – 0.60; *p* = 0.04) from baseline to post-intervention. In contrast, text-message specific smartphone-based mHealth interventions only led to significant increases in steps/d (SMD 0.34; 95% CI 0.02 – 0.66; *p* = 0.04; Feter et al., 2019). Lastly, Romeo et al. (2019) conducted a meta-analysis to determine the effectiveness of smartphone based mHealth apps for increasing PA in adults using randomized controlled trials (RCT) only. Results demonstrated that smartphone apps produced a nonsignificant (*p* = 0.19) increase in participant average daily step count in comparison to control conditions, with a mean difference of 476.75 steps/d (95% CI -229.57 – 1183.07) between groups. Sensitivity analyses demonstrated that PA interventions using smartphone apps for less than three months (versus greater than three months) were more effective and significantly increased PA by 2074.96 steps/d (95% CI 606.80 – 3543.11, SMD 0.56, 95% CI 0.16 – 0.97; *p* = 0.01). Apps that targeted PA in isolation (versus combined interventions of PA and diet) were more effective and significantly increased step count by 716.86 steps/d (95% CI 38.37 – 1395.86; *p* = 0.04, SMD 0.31, 95% CI 0.07 – 0.00; *p* = 0.01). Differences in step count were not significantly different between general adult populations and those with specific health conditions (Romeo et al., 2019).

2.2.3 ***Attrition.*** Attrition in mHealth interventions is a measure of disengagement and is comprised of dropout and nonusage attrition (Eysenbach, 2005). Dropout attrition concerns intervention retention and is characterized by participants not returning to complete follow-up measurement (Eysenbach, 2005). Dropout attrition decreases the power of a study and complicates the interpretation of results because there is no knowledge of the intervention effect in those that did not provide follow-up data (E. Murray et al., 2013). Notably, the meta-analysis by Laranjo et al. (2021) of RCT evidence on the effectiveness of mHealth PA interventions found that dropout attrition was less than 10% and a significant predictor of intervention effectiveness. However, mHealth interventions using smartphone apps and activity trackers were included in the meta-analysis (Laranjo et al., 2021). Slightly higher rates of dropout attrition were found in PA studies included in a meta-analysis of attrition in app-based mHealth interventions for chronic disease (Meyerowitz-Katz et al., 2020). Hales, Turner-McGrievy, Wilcox et al. (2016) conducted a two-armed RCT to test the efficacy of the *Social POD* app which targeted social support, dietary self-monitoring, PA, and weight among overweight and obese adults. Compared to the *Calorie Counter by Fat Secret* app (FatSecret, n.d.) used by the control group, the *Social POD* app included social networks, regular notifications, and FIs in the form of points that were redeemable for prizes (Hales, Turner-McGrievy, Fahim, et al., 2016). Dropout attrition was identical (12%) in both experimental (n = 3) and control (n =3) conditions out of the 51 participants who began the study (Hales, Turner-McGrievy, Wilcox, et al., 2016). Spring et al. (2018) examined whether the *Make Better Choices 2* app (designed for the study) could sustainably improve diet and PA using FIs and remote coaching in a 9-month three-arm prospective RCT. The app was used to deliver two intervention conditions of diet and PA that was compared with a control condition which coached participants to perform three daily relaxation exercises (progressive muscle relaxation, mindfulness meditation, and self-hypnosis). The intervention conditions targeted MVPA simultaneously with or sequentially after diet (fruit and vegetable intake) and activity risk behaviour (sedentary leisure time). In the sequential condition, the PA interface of the app was unavailable until week 7. Dropout attrition rates in the simultaneous, sequential, and control conditions were 19%, 17.8%, and 15.9%, respectively (Spring et al., 2018).

Nonusage attrition refers to intervention adherence and describes the propensity of study participants to either not use or discontinue to use an mHealth intervention. Nonusage attrition leads to an underestimate of the potential efficacy of an intervention given that maximal health benefits are associated with adherence to intended use (i.e., following prescribed recommendations of the intervention; Sieverink et al., 2017; E. Murray et al., 2013). Relative to dropout attrition, RCT evidence of nonusage attrition in mHealth PA interventions is lacking and was only evaluated in one observational study included the meta-analysis by Meyerowitz-Katz et al. (2020). In this observational study, Guertler et al. (2015) investigated nonusage attrition in the 10,000 Steps program, a free PA promotion initiative delivered on the internet and as a smartphone app (Government of Australia, n.d.). Three participant subgroups were defined by the platform used to log steps: web-only users who only utilized the website, app-only users who only utilized the smartphone app, and web-and-app users who utilized both the website and smartphone app. Nonusage attrition was defined as the duration of program use (days) before a user did not log PA for at least 14-days. Nonusage attrition did not occur for 25% of web and app-only users until 41 and 43 days of program use, respectively. Comparatively, nonusage attrition did not occur for 25% of web-and-app users until 56 days of program use. Univariate analysis indicated that risk of nonusage attrition was significantly reduced in app only (HR 0.86, SE 0.03, 95% CI 0.58 – 0.68; $p < 0.001$) and web-and-app (HR 0.63, SE 0.03, 95% CI 0.81 – 0.93; $p < 0.001$) users relative to web-only users (Guertler et al., 2015). After the search for eligible studies in the meta-analysis by Meyerowitz-Katz et al. (2020) concluded (i.e., June 2019), Edney et al. (2019) published a 100-day secondary analysis of an RCT that examined nonusage attrition between a gamified and basic version of the smartphone-based app, *Active Team*. Both versions encouraged participants to take 10,000 steps per day and sent daily reminders to self-monitor steps. The gamified version utilized gamification and included additional features such as a Facebook-style newsfeed, PA challenges, a leaderboard, and unlockable virtual gifts (Deterding et al., 2011; Edney et al., 2017). Nonusage attrition was defined to occur when users ceased to access the app for 30 consecutive days or more which occurred for 31.9% and 39.4% of the gamified and basic groups, respectively. There were no significant between-group differences in time to nonusage attrition (p = 0.17; Edney et al., 2019).

2.2.4 ***Engagement.*** Increased engagement in web-based interventions has been shown to reduce dropout (Couper et al., 2010) and nonusage attrition (Kelders et al., 2012). Engagement in mHealth interventions has been defined as the “extent (e.g., amount, frequency, duration, and depth) of usage and subjective experience of users characterized by attention, interest and affect” (Perski et al., 2017). Engagement is necessary for the effectiveness of mHealth interventions (Perski et al., 2017; Yardley et al., 2016) and is not synonymous with ‘adherence’ which is defined as the proportion of participants who use an intervention as it is intended to be used (Kelders et al., 2012). Maintained engagement in mHealth interventions is especially difficult without human-to-human interaction and can lead to increased attrition (L. F. M. Kohl et al., 2013). Despite its importance, only one meta-analysis has examined the association between levels of engagement with DHIs and PA which included, but was not specific to, mHealth interventions (McLaughlin et al., 2021). McLaughlin et al. (2021) conducted a meta-analysis with the primary objective to investigate the direction and strength of the association between DHI engagement (measured by extent of usage and subjective experience) and PA. Explorations into whether the direction of association between DHI engagement and PA varied by type of engagement measure (i.e., extent of usage and subjective experience) was a secondary objective. Under the definition of engagement conceptualized by Perski et al. (2017), extent of usage was objectively measured by the number of activities completed and logins along with total time spent on the DHI. Subjective experience was assessed by measures of attention, interest, and affect such as enjoyment, satisfaction, user experience, and usability. For the primary objective, the pooled estimate of the standardized regression coefficient (SRC) indicated a small but significant positive relationship with extent of usage and PA (0.08; 95% CI 0.01 – 0.14; $p = 0.02$; SD 0.11; I2 = 77%) in 11 studies. However, subjective experience could not be examined due to considerable heterogeneity and the small number of studies ($n = 3$) reporting it as an outcome. Vote counting was implemented to address the secondary objective which indicated that most associations (15 of 26 studies) supported the study hypothesis of increased engagement being associated with higher PA. For type of engagement measure, the study hypothesis was consistently supported for subjective experience (two of three), activities completed (five of eight), and logins (six of 10). A positive association was not consistently found for time (n = five associations) as two studies reported inconclusive findings and one rejected the study hypothesis. Two studies of mHealth interventions that measured extent of usage exceeded the pooled estimate of the SRC (0.08) found in the primary analysis which corresponded to 0.187 (Edney et al., 2019) and 0.125 (Marquet et al., 2018), respectively (McLaughlin et al., 2021).

In addition to nonusage attrition, Edney et al. (2019) examined engagement in the 100-day secondary analysis of an RCT which compared a gamified and basic version of the smartphone-based app, *Active Team*. Engagement was measured by total app use (number of times app features were used during the 100-day intervention period) along with daily and monthly active users (number and percentage of gamified and basic app users who accessed the app daily and at least once every 30 days, respectively). In addition, PA was also assessed among the most highly engaged users, defined as superusers (users in the top quartile of total app use). Results indicated there was a weak but significant total app use-by-time interaction effect for MVPA measured by accelerometer ($F_{1,272}$ = 4.5; p = 0.04) and self-report ($F_{1,304}$ = 6.56; p = 0.01), where higher feature use was associated with increased PA at 3-month follow-up. Furthermore, there was a significant group by time interaction, where superusers completed 28.2 more min/d of MVPA than regular users at 3-month follow-up (SE = 9.5, 95% CI 9.4 – 46.9, $F_{1,272}$ = 4.76; p = 0.03; Edney et al., 2019). Marquet et al. (2018) performed a cross-sectional study of new and existing users (n = 74) of the smartphone-based app *Pokémon Go*. *Pokémon Go* is defined as an augmented reality geocaching exergame (Baranowski, 2017) and while PA is not a direct aim of the game, it is a mechanism through which players can progress. For instance, some features are only unlocked when certain walking thresholds are met and key locations in the game (*PokéStops and Pokémon Gyms*) require players to be physically proximate to be used (Ninantic Inc., n.d.). Engagement outcomes were measured by time through total playing minutes and number of playing episodes per day. For participants who self-identified as *Pokémon Go* players (n = 47), a significant partial correlation was found (r = 0.176; $p < 0.05$) between total playing minutes per day and number of steps measured by accelerometry and ecological momentary assessment (EMA). Three active (walking, jogging, bicycling) playing episodes per day were associated with an increase of 1526 steps compared to not playing or playing without being active (95% CI 329.32 – 2723.9; $p = 0.013$; Marquet et al., 2018).

2.3 *Incentive-based Interventions and Physical Activity*

2.3.1 ***Incentive-based Digital Health Interventions and Engagement.*** Experimental and quasi-experimental evidence indicates that FIs can increase engagement and PA when incorporated into web-based (J. M. Murray et al., 2019; Omran et al., 2018; West et al., 2020) and mHealth interventions (Mitchell et al., 2018; Mitchell, Lau, et al., 2020). Furthermore, the effects appear to differ by baseline level of PA (Mitchell et al., 2018; Mitchell, Lau, et al., 2020) and may be mediated by integrated regulation (J. M. Murray et al., 2019). Omran et al. (2018) conducted an 11-week RCT on a web-based walking intervention to examine the effect of providing FIs for self-regulatory behaviours (i.e., self-monitoring and action planning) in inactive office employees. Participants were randomized into control (intervention only) and FI (intervention plus CAD $5.00 electronic-gift card delivered weekly for completing action plans over 4-weeks) conditions. Engagement was measured by completion of action plans using an action planning tool which aimed to help participants implement short, planned walks into their daily routines and encourage achievement of a daily PA target (i.e., 2000 steps/d over baseline step count). Physical activity was measured by step count using a pedometer and self-reported daily on the website. A large effect size in favor of the FI condition was observed for the average number of action plans completed during the incentive period (Cohen's *d* = 1.01) which persisted after FIs were withdrawn (Cohen's *d* = 1.00). A large effect size (Cohen's *d* = 0.62) was found for change in average daily step count between baseline and the post-incentive period for the FI condition (mean: $\bar{x}$ = 1793, SD = 2408.72). The control condition demonstrated a small effect size (Cohen's *d* = 0.24) for change in average daily step count between baseline and the post-incentive period ($\bar{x}$ = 686, SD = 2887.62; Omran et al., 2018).

West et al. (2020) performed a 6-month RCT on a web-based behavioural weight control program in overweight and obese adults. Participants were randomized into a 24-session online group-based intervention with weekly synchronous chat sessions (internet-only) or the same program providing weekly FIs (Amazon electronic gift card) for self-monitoring body weight, daily dietary intake, and achieving targeted weight loss at 2- and 6-months (internet plus FI). Participants were asked to provide daily updates on the study website specifying whether they met their caloric intake goal, how many minutes of MVPA they completed, number of steps taken, and if they weighed themselves (if so, to report their weight). These self-reported website updates were used to measure treatment engagement in both conditions and inform weekly payouts for participants in the internet plus FI condition. During the first 2-months of FI provision, significant increases in self-reported PA goal attainment (number of weeks ≥ 200 min/wk of MVPA and number of days ≥ 10,000 steps/d) were observed in the internet plus FI condition (MVPA: $\bar{x}$ = 3.3, SD = 2.9; $p < 0.0001$ and steps: $\bar{x}$ = 15.1, SD = 14.7; $p < 0.0001$) compared to the internet-only condition. Self-reported PA volume (daily minutes of MVPA and number of steps) in the internet plus FI condition was also significantly greater (MVPA: $\bar{x}$ = 200, SD = 127; *p*

<0.0001 and steps: $\bar{x} = 7806$, SD = 2659; $p = 0.01$) than the internet-only condition. After FI removal, self-reported PA goal attainment was significantly higher in the internet plus FI condition (MVPA: $\bar{x} = 5.5$, SD = 5.8; $p < 0.0001$ and steps: $\bar{x} = 27.0$, SD = 29.3; $p = 0.0004$) compared to the internet-only condition. In terms of volume however, only self-reported MVPA demonstrated a significant increase in the internet plus FI condition ($\bar{x} = 191$, SD = 148; $p = 0.003$) compared to internet-only condition. From baseline to 6-month follow-up, significant differences were observed for self-reported PA goal attainment (MVPA: $\bar{x} = 9$, SD = 8; $p < 0.0001$ and steps: $\bar{x} = 42$, SD = 41; $p = 0.0003$) and self-reported PA volume (MVPA: $\bar{x} = 193$, SD = 131; $p < 0.0001$ and steps: $\bar{x} = 8057$, SD = 2672; $p = 0.04$) in the internet plus FI condition relative to the internet-only condition (West et al., 2020).

J. M. Murray et al. (2019) conducted a process analysis on the physical activity loyalty (PAL) scheme cluster RCT to determine if engagement among intervention components predicted PA and psychosocial mediators (i.e., intrinsic motivation) of behaviour change 6-months post-baseline. The PAL scheme was a 6-month multicomponent web-based intervention targeting workplace PA. Financial incentives were incorporated in an evidence-based behaviour change program informed by learning theory, social cognitive theory and SDT that included self-regulation techniques (Bandura et al., 1999; Deci et al., 1999; Hunter et al., 2016; Johnston, 2016; Michie et al., 2014). Engagement over the 6-month intervention period was measured as the percentage of days during which participants walked for ≥ 10 minutes, the percentage of weeks during which participants logged onto the website at least once, and the percentage of earned points redeemed (worth £0.03 for a maximum of 30 minutes walking/d). Engagement with different aspects of the website was assessed as the frequency of hits on each intervention component for every 10 days a participant accessed the website and the total number of components accessed on the website at least once (range zero to six). The six intervention components participants could access on the website included monitoring and feedback, rewards, maps, health information, health information specific to PA, and discussion forums. Physical activity was measured by steps/d using sealed pedometers. Self-reported psychosocial mediators of planning, self-determined motivation, habit, recovery and maintenance self-efficacy, outcome satisfaction, social norms, and workplace norms were collected at baseline and 6-months. Engagement variables that significantly predicted steps/d at 6-months in univariable analysis were included in a multivariable model which showed the frequency of hits on the monitoring and feedback component of the website (b = 50.2, SE = 24.5; p =0.04) and the percentage of earned points redeemed (b = 9.1, SE = 3.3; p = 0.005) were positively related to steps/d at 6-months. Using a multivariable model, engagement variables that significantly predicted 6-month integrated regulation were the frequency of hits on the monitoring and feedback component of the website (b = 0.03; SE=0.01; p = 0.02) and percentage of days which participants walked for at least 10 minutes (b = 0.008; SE = 0.002; p < 0.001) which positively related to steps/d at 6-months. Notably, engagement with the rewards component of the intervention (reward redemption) was not related to levels of identified regulation, integrated regulation, and intrinsic motivation (J. M. Murray et al., 2019).

Mitchell et al. (2018) investigated whether the *Carrot Rewards* app, a multicomponent mHealth intervention that included goal setting, graded tasks, biofeedback, and FIs for daily step goal achievement, could increase PA in two Canadian provinces (i.e., BC and

NL). The 12-week single group pre-post quasi-experiment (QE) included 32,229 participants who were enrolled in the “Steps” walking program (and therefor eligible to receive FIs worth CAN $0.04 for daily step goal achievement) and had valid baseline step count data (i.e., ≥ 5 days during 14-day baseline period of 1000 ≥ 40,000 steps/d). Participants were categorized by baseline mean steps/d as physically inactive or physically active (<5000 and ≥ 5000, respectively). Participant engagement was dichotomized into categories of “high” or “low” based on the median percentage of days when a “Step Up Challenge” was accepted. “Step Up Challenges” provided participants with FIs worth CAN $0.40 for reaching daily step goals ≥ 10 non-consecutive times within a 14-day period after being enrolled in the “Steps” program for at least two weeks. Mixed-effects models were conducted for data analysis and local effect sizes were calculated using Cohen f^2, with $f^2 \geq 0.02$, $f^2 \geq 0.15$, and $f^2 \geq 0.35$ representing small, medium and large effect sizes, respectively (Selya et al., 2012). Results indicated significant increases in mean daily step count when baseline was compared with each study week ($p < 0.001$). From baseline to week 12, participant step count increased by 115.70 (95% CI 74.59 – 156.81; $p < 0.001$) with a small effect size (Cohen $f^2 = 0.0059$). Participants with “high” engagement in BC and NL increased step count by 738.70 (95% CI 673.81 – 803.54; $p < 0.001$) and 346.00 (95% CI 239.26 – 452.74; $p < 0.001$) steps/d, respectively. Participants who were physically inactive and with “high” engagement averaged an increase of 1224.66 steps/d (95% CI 1160.69 – 1288.63; $p < 0.001$; Mitchell et al., 2018).

In a follow up study, Mitchell, Lau, et al. (2020), examined the impact of *Carrot Rewards* on PA over 12-months in BC and NL. Participants (n=39,113) were categorized into four engagement groups (‘Limited’: 1-11 weeks, ‘Occasional’: 12-23 weeks, ‘Regular’: 24-51 weeks, and ‘Committed’: 52 weeks) based on the number of weeks with four or more days of valid step count data (i.e., $1000 \geq 40{,}000$ steps/d). Participants were classified by baseline mean steps/d as physically inactive and physically active (< 5000 and ≥ 5000 steps/d, respectively). Mixed-effects models were used for data analysis and local effect sizes were calculated using Cohen f^2. Findings indicated that differences between baseline and average of the last two recorded weeks were statistically significant ($p < 0.0001$) for all sub-group analyses (engagement groups and PA status within engagement group). Average daily step count significantly increased in ‘Regular’ (least-square means [LSM] 448.8, 95% CI 407.9 – 489.7) and ‘Committed’ (LSM 884.6, 95% CI 824.8 – 944.4) participants but significantly decreased in ‘Limited’ (LSM -392.3, 95% CI -439.9 – [-344.7]) and ‘Occasional’ (LSM -473.2, 95% CI -527.4 – [-418.9]) participants. Small effect sizes were observed in ‘Committed’ and ‘Occasional’ participants (Cohen’s $f^2 = 0.0563$ and 0.0211, respectively). The greatest differences in mean daily step count were observed in physically inactive ‘Regular’ (LSM 1215, 95% CI 1163 - 1266) and ‘Committed’ (LSM 1821, 95% CI 1739 - 1902) participants with medium effect sizes (Cohens $f^2 = 0.1617$ and 0.3140, respectively). Furthermore, physically inactive participants in lower engagement groups demonstrated greater differences in mean daily step count (‘Limited’ = 388.6 and ‘Occasional’ = 435.5) than those categorized as physically active (‘Limited’ = -957.9 and ‘Occasional’ = -1141; Mitchell, Lau, et al., 2020).

2.3.2 ***Incentive Design in Smartphone-Based mHealth Interventions.*** Experimental and quasi-experimental evidence indicate that the effectiveness of smartphone-based mHealth interventions that utilize FIs to increase PA are influenced by incentive design (Patel et al., 2018; Patel, Asch, Rosin, Small, Bellamy, Eberbach, et al., 2016; Patel, Asch, Rosin, Small, Bellamy, Heuer, et al., 2016; Patel, Volpp, Rosin, Bellamy, Small, Fletcher, et al., 2016; Pearson et al., 2020). Patel, Asch, Rosin, Small, Bellamy, Eberbach, et al. (2016) compared the effectiveness individual versus team-based FIs to increase PA using the *Moves* app in a 26-week four-armed RCT (Moves, 2018). All participants used a smartphone to track activity and received daily feedback on performance for achieving a goal of 7000 steps/d during the intervention and follow-up periods that each lasted 13-weeks. In the three FI arms, drawings were conducted that selected one winning team every other day during the 13-week intervention. Participants on a winning team were eligible to receive US $50 if the goal was met individually (individual FI), US $50 if all four team members met the goal (team FI), or US $20 if the goal was met individually and US $10 for each of three teammates that also met the goal (combined FI). Compared to the control group during the intervention, the mean proportion achieving the step goal was only significantly greater for the combined FI (difference: 0.17, 95% CI 0.07 – 0.28; $p < 0.001$). The combined FI arm achieved the goal at significantly greater rates than the team FI (difference: 0.18, 95% CI 0.08 – 0.28; $p < 0.001$) but not the individual FI (difference: 0.10, 95% CI -0.001 – 0.19; $p = 0.05$). Only the combined FI had significantly greater mean daily steps than the control group (difference: 1446 steps/d, 95% CI 448 – 2444; $p \leq 0.005$). Goal achievement decreased during the follow-up period after FI removal and there were no significant differences between arms (Patel, Asch, Rosin, Small, Bellamy, Eberbach, et al., 2016).

Patel, Asch, Rosin, Small, Bellamy, Heuer, et al. (2016) tested the effectiveness of three FI framing methods to increase PA among overweight and obese adults using the *Moves* app in a 26-week four-armed RCT (Moves, 2018). All participants used a smartphone to track activity and received daily performance feedback for achieving a goal of 7000 steps/d during the 13-week intervention and follow-up periods. The three FI arms were gain-framed (US $1.40 given on each day of goal achievement), lottery-based (daily eligibility [approximate expected value of US $1.40] based on goal achievement), and loss-framed (US $42 allocated upfront monthly and US $1.40 removed each day the goal was not achieved). In adjusted analysis, only the loss-framed FI group had a significantly greater mean proportion of days achieving the goal than controls (adjusted difference: 0.16, 95% CI 0.06 – 0.26; $p = 0.001$), but was not significant different for mean daily steps (adjusted difference: 861 steps/d, 95% CI 24 – 1746; $p = 0.056$). During follow-up after FI removal, daily steps decreased for all groups with no significant differences (Patel, Asch, Rosin, Small, Bellamy, Heuer, et al., 2016).

Patel, Volpp, Rosin, Bellamy, Small, Fletcher, et al. (2016) examined the effectiveness of different combination of social comparison feedback and FIs to increase PA using the *Moves* app in a 26-week 2 x 2 factorial RCT (Moves, 2018). All participants used a smartphone to track activity and received daily performance feedback for achieving a goal of 7000 steps/d along with social comparision feedback during the 13-week intervention and follow-up periods. Two hundred eighty-eight participants formed teams of 4 members and were randomly assigned to receive 1 of 2 types of team-based performance feedback either with or without FIs. In 2 arms, participants received weekly feedback on team performance (average daily steps per team member) and no FI. In 1 arm, each team was informed how their weekly average step count compared to the 50th percentile (median) from the same arm (above or below, and average step count). In the other arm, each team was told how their weekly average step count compared to the 75th percentile (top quartile). In the 2 FI arms, teams received the same feedback of how their weekly average step count compared to either the 50th or 75th percentile and were entered into a weekly lottery. The daily approximate expected value per participant was US $1.40 who were only eligible to receive the FI if their average step count per day per team member during the week prior was 7000 steps or higher. Results indicated that mean proportion of goal achievement was only significantly greater for the 50th percentile with FI group compared to the 75th percentile without FI during the intervention period (difference: 0.18, 95% CI 0.04 – 0.32; $p = 0.012$). During the follow-up period after FI removal there were no significant differences between any group in mean proportion of goal achievement or daily steps (Patel, Volpp, Rosin, Bellamy, Small, Fletcher, et al., 2016).

Patel et al. (2018) tested the effectiveness of varying lottery-based FIs to increase PA among overweight and obese adults using the *Moves* smartphone app in a 26-week 4-armed RCT (Moves, 2018). All participants used a smartphone to track activity and received daily performance feedback for achieving a goal of 7000 steps/d during the 13-week intervention and follow-up periods. The 3 lottery-based FI arms were higher frequency, smaller reward (1 in 4 chance of winning US $5), jackpot (1 in 400 chance of winning US $500), or combined (18% chance of winning US $5 and 1% chance of winning US $50) and contingent on goal achievement from the day prior. In adjusted models, only participants in the combined lottery arm had significantly greater odds of goal achievement than those in the control group (odds ratio [OR], 3.00, 95% CI 1.28 – 7.02; $p = 0.012$). Notably, the weekly trend analysis indicated a significant decline in proportion of goal achievement among jackpot arm participants (-0.011 per week, 95% CI -0.017 – [-0.005]; $p < 0.001$], accounting for a 0.13 decrease compared to the control group during the intervention period. Mean proportion of goal achievement and daily steps (unajusted and adjusted) declined during the follow-up period with no significant differences between arms (Patel et al., 2018).

Pearson et al. (2020) examined the impact of adding team-based FIs called "Step Together Challenges" to the *Carrot Rewards* app in a 24-week QE (retrospective matched pairs design) in three Canadian provinces, BC, NL, and ON. The experimental group included participants who used the "Step Together Challenge" feature for the first time between March 19 and April 16, 2018. "Step Together Challenges" enabled participants to earn a $0.40 CAD bonus for collaboratively reaching ≥10 individual daily step goals in a 7-day period with a friend. Control participants were selected from a cohort of *Carrot Reward* users who had enabled the "Steps" walking program but did not engage in a "Step Together Challenge" during the study period. Experimental and control participants were matched on age (± 1 year), gender, province, and baseline step count (± 500 steps/d). Controlling for pre-intervention mean daily step count, analysis of covariance (ANCOVA) showed a significant difference in intervention mean daily step count in the experimental group compared to the control group (F [1, 61,167], $p < 0.0001$; $\eta_p^2 = 0.024$). The estimated marginal mean group difference was 537 steps/d, or 3759 steps/wk. Linear regression suggested a dose-response relationship between the number of "Step Together Challenges" completed and intervention mean daily step count

(F [1, 14] = 35.834, $p < 0.0001$, adjusted $R^2 = 0.699$). Participants' mean daily step count increased 196.80 (unstandardized beta coefficient) for each new "Step Together Challenge" completed, on average (Pearson et al., 2020).

2.4 *Incentive Removal and Sustained Physical Activity*

2.4.1 ***Contemporary Evidence.*** Two recent meta-analyses indicate that FI utilization in interventions can increase PA which is maintained in the post-incentive period. Luong et al. (2020), found moderate-quality evidence at end of intervention that FIs increased walking behaviour measured by daily steps (SMD 0.25, 95% CI 0.13-0.36, $p < 0.01$; $I^2 = 55\%$) and leisure time PA measured by gym attendance (SMD 0.46, 95% CI 0.28-0.63, $p < 0.0001$; $I^2 = 84\%$) corresponding to small and moderate effects, respectively. At longest follow-up, moderate-to-high quality evidence indicated a small effect of FIs to sustain increases in walking behaviour (SMD 0.11, 95% CI 0.00-0.22, $p = .07$; $I^2 = 39\%$) and leisure time PA (SMD 0.10, 95% CI 0.02-0.19, $p = .0154$; $I^2 = 3.3\%$). Studies with the greatest difference from the overall SMD (0.11) for walking behaviour at longest follow-up used FI design features that were loss-framed cash (SMD 0.52; Chokshi et al., 2018), cash, donation, or a combination (cash and donation) in older adults (≥ 65 years [SMD 0.35]; Harkins et al., 2017), vouchers for creating and completing a PA action plan (SMD 0.32; Omran et al., 2018), and lotteries for specific goods or services among adults over the age of 55 (SMD 0.93; Petry et al., 2013). Studies with smaller differences from the overall SMD (0.11) for walking behaviour at longest follow-up used FI design features that provided cash for individuals, teams or a combination (individual and team) of both (SMD 0.15; Patel, Asch, Rosin, Small, Bellamy, Eberbach, et al., 2016) and with social comparison feedback (SMD 0.15; Patel, Volpp, Rosin, Bellamy, Small, Fletcher, et al., 2016). The only studies on leisure time PA that exceeded the overall SMD of 0.10 post-intervention used cash-specific FIs of at least US $116.63 (SMD 0.22; Acland & Levy, 2015 and SMD 0.46; Charness & Gneezy, 2008) and vouchers of US $254.27 (SMD 0.26; Condliffe et al., 2017). In absolute terms, a slightly larger mean difference (MD) in daily steps (754 steps/d) was observed than by Mitchell, Orstad, et al. (2020; 607 steps/d) at end of intervention and a smaller difference at longest follow-up (MD = 459 steps/d vs. MD = 376 steps/d, respectively). Given that secondary follow-up time points ranged from four to 104 weeks, findings of sustained effects should be interpreted with caution. Effects at longest follow-up were weaker than the moderate effects observed at the end of intervention and only 10/31 trials measured PA > 6 months after the FI period (Luong et al., 2020).

Mitchell, Orstad, et al. (2020) extended on findings from Luong et al. (2020) by reporting pooled MD in daily step count by study and participant characteristics using a subgroup meta-analysis. Results from the primary meta-analysis found that FIs increased mean daily step count during the intervention (pooled MD 607.1 steps/d, 95% CI 422.1 – 792.1) and post-intervention (pooled MD 513.8 steps/d, 95% CI 312.7 – 714.9) period. Design features of studies that exceeded the overall pooled MD (513.8 steps/d) in the post-intervention period used cash for individuals (MD 1026 steps/d; Harkins et al., 2017) or donations (MD 1099 steps/d; Harkins et al., 2017), cash lotteries (MD 3015 steps/d; Kullgren et al., 2014) or cash lotteries combined with social support (MD 1833 steps/d; Kullgren et al., 2014), loss-framed cash (MD 526 steps/d; Patel, Asch, Rosin, Small, Bellamy, Heuer, et al., 2016), cash for individuals and teams combined (MD 1077 steps/d; Patel, Asch, Rosin, Small, Bellamy, Eberbach, et al., 2016), and lotteries redeemable for specific goods or services (MD 2499 steps/d; Petry et al., 2013). Heterogeneity was high during the intervention (I^2 = 80.8, $p < 0.0001$, Q = 114.5) and post-intervention (I^2 = 85.1, $p < 0.0001$, Q = 120.8) period but was an expected result from data pooling using multicomponent behaviour interventions. However, heterogeneity was moderate to low in the subgroup meta-analysis indicating the differences may have been accounted for by study and participant characteristics (Higgins & Thompson, 2002). In the subgroup meta-analysis, the greatest differences during the incentive period were detected in studies using wearable tracking devices (versus smartphones; 834 steps/d), larger FIs (above versus below median, US $1.40; 354 steps/d), less active (versus non-specific; 474 steps/d) and older adults (versus non-specific; 358 steps/d). During the post-intervention period, the greatest differences were observed in studies using wearable tracking devices (versus smartphones; 620 steps/d), larger FIs (above versus below median, US $1.40; 620 steps/d), and overweight or obese adults (versus non-specific; 411 steps/d). Studies with longer intervention period (> 23 weeks) led to larger post-intervention effects (versus interventions lasting 12 – 23 weeks; 467 steps/d; Mitchell, Orstad, et al., 2020).

2.4.2 ***Relevant Behaviour Change Theories.*** The effect of FIs on PA can be described using theories of motivation from psychological and BE literature (Promberger & Marteau, 2013). Grounded in psychology, SDT suggests that motivation is reflected along an internalization continuum which represents the degree to which a behaviour has been self-integrated (R. M. Ryan & Deci, 2000). At each end of the internalization continuum are amotivation (i.e., lack of motivation) and intrinsic regulation (i.e., intrinsic motivation), where the former reflects a lack of intention to act and the latter is autonomous, characterized by participation for the self-rewarding nature of the behaviour. Four types of extrinsic motivation separate amotivation from intrinsic regulation within the internalization continuum which vary in terms of self-integration. External regulation and introjected regulation manifest controlling internalizations that motivate behaviour through a desire to appease others, avoid negative feelings or maintain conditional self-worth. Identified regulation and integrated regulation are more autonomous extrinsic motives, where the former is characterized by participation regulated by goal values or importance of behavioural outcomes and the latter is represented by congruency between behavioural regulation and personally endorsed values, goals and needs that are already part of the self (R. M. Ryan & Deci, 2000). Position along this continuum is determined by the extent to which basic psychological needs for competence, autonomy and relatedness are satisfied by social contexts (R. M. Ryan & Deci, 2000). Indeed, there is a consistent and positive relationship between more autonomous forms of motivation and PA behaviour. Identified regulation has been shown to predict initial and short-term participation more strongly than intrinsic motivation whereas intrinsic motivation is more predictive of long-term adherence (Teixeira et al., 2012). Research on SDT has shown that intrinsic motivation measured by behavioural persistence (i.e., time spent) declines in response to extrinsic rewards which has been defined as the undermining effect (Deci et al., 1999). However, most research on the undermining effect has measured intrinsic motivation through time spent on simple and enjoyable tasks such as puzzles for which initial levels of behaviour and intrinsic motivation are high (Deci et al., 1999; Promberger & Marteau, 2013). No evidence of an undermining effect has been found when FIs are provided for health behaviour in which initial levels of intrinsic motivation and participation are low (Promberger & Marteau, 2013; Deci & Ryan, 2002). Cognitive evaluation theory (CET) is a subtheory of SDT which predicts that providing FIs for attainable and confidence-promoting goals may increase intrinsic motivation by mediating perceived competence and autonomy (R. M. Ryan & Deci, 2000). Unfortunately, in the meta-analysis by Mitchell, Orstad, et al. (2020), no studies measuring self-determined motivation over time were found and predictions from CET relating to FI removal and sustained PA could not be tested (Mitchell, Orstad, et al., 2020).

Standard economic theories employ a behavioural model which typically disregard psychological factors (Frey & Benz, 2005) and assumes that individuals are utility maximizers (i.e., rational, self-controlled, and self-interested; Camerer & Loewenstein, 2003). Assumptions that individuals are rational, self-controlled, and self-interested have been systematically challenged by findings in BE (Weibel et al., 2014). The aim of BE is to stepwise modify the conventional assumptions of standard economic theories to build a more practical psychological-empirical foundation of standard economic models (Rabin, 2002). As a result, BE has developed a more comprehensive understanding of human motivation than standard economic theories (Weibel et al., 2014). Behavioural economics theorizes that extrinsic and intrinsic motivation cannot be seen as an additive phenomenon (Weibel et al., 2014), but instead interact in a predictable way (Frey & Jegen, 2001). Motivation crowding theory (MCT), a subtheory of BE (Frey, 2017), builds on CET and identifies circumstances that undermine (“crowd-out”) or strengthen (“crowd-in”) intrinsic motivation in response to FIs (Promberger & Marteau, 2013). Intrinsic motivation can be crowded-out through decreased self-determination and self-esteem if individuals perceive an intervention to be controlling. Alternatively, intrinsic motivation can be crowded-in if individuals perceive an intervention to be supportive which augments self-esteem and self-determination (Frey & Jegen, 2001).

Behavioural economics and SDT may account for how FIs can increase PA that is sustained after withdrawal when considered under the transtheoretical model of health behaviour change (Moschetti, 2013). The transtheoretical model suggests that health behaviour change involves progression through six stages of change: precontemplation, contemplation, preparation, action, maintenance, and termination (Prochaska & Velicer, 1997). In the precontemplation stage, individuals do not intend to initiate a healthy behaviour in the foreseeable future (i.e., six months) whereas individuals in the contemplation stage do intend to start a healthy behaviour within the next six months. Preparation is the stage in which individuals intend to begin a healthy behaviour in the immediate future (i.e., one month), while action is the stage in which individuals have initiated a healthy behaviour within the past six months. In the maintenance stage, individuals work to prevent relapse into prior unhealthy behaviour patterns and continue to participate in the adopted health behaviour for six months to five years. Lastly, termination is the stage in which individuals have zero temptation to relapse into prior unhealthy behaviour patterns and have complete self-efficacy to continually participate in the adopted health behaviour. Notably, only about 15% of individuals in the maintenance stage relapse to the precontemplation stage in terms of PA (Prochaska & Velicer, 1997). As individuals move through the stages of the transtheoretical model, FIs can be used to ‘nudge’ them from precontemplation into action (Moschetti, 2013). If the design of a FI intervention satisfies the three basic psychological needs (i.e., autonomy, competence, and relatedness), individuals who reach the maintenance stage may have internalized the newly acquired health behaviour which could be sustained (R. M. Ryan & Deci, 2000) while FI are eventually withdrawn (Moschetti, 2013).

2.4.3 ***Contradictions of 'Undermined' or 'Crowded-out' Intrinsic Motivation.*** Several studies have demonstrated evidence to contradict 'undermined' or 'crowded-out' intrinsic motivation for PA from FIs (Charness & Gneezy, 2011; Kramer et al., 2020; J. M. Murray et al., 2020; Pope & Harvey, 2015). Charness & Gneezy (2008) examined habit formation (HF; Becker, 1988) and MCT (Frey & Jegen, 2001) in two sequential experiments to investigate the impact of FIs on PA post-intervention using campus fitness center attendance in university students. According to HF, habits are thought to be harmful or beneficial depending on the extent to which they decrease or increase future utility. Given that the marginal utility of current consumption correlates with past consumption, changes in the present that have small short-term effects may have increasingly large effects in the future (Becker, 1992). Charness & Gneezy (2008) hypothesized that if PA is a form of habitual behaviour, future utility may be increased by providing FIs for regular participation. If marginal utility of current consumption positively correlates to past consumption, periods of FI provision may induce people to participate in PA more regularly in the future. The studies tested whether FIs would reduce attendance in accordance with the crowding-out effect postulated by MCT or increase gym attendance by inducing HF. Participants (n=120) in the first study were randomized into control (n=40) and FI conditions where $25 was initially granted for ≥ 1 visit to the fitness center during the following week. Participants in the FI conditions (n=80) were randomized into one of two groups; for half of them this was the end of the experiment while the other half was promised an additional $100 for attendance ≥ 8 visits over the next four weeks. In the second study, all participants (n=168) were paid $175 in installments to attend three biometric tests. The first FI condition required participants to attend the gym ≥ 1 whereas the second FI condition required ≥ 8 visits over 4-weeks, respectively. In both studies, evidence supported HF with significant increases gym attendance after FI removal, particularly among participants without baseline regular attendance. However, results were also consistent with crowding-in in accordance with MCT (Frey & Jegen, 2001). Partial support for crowding-out was demonstrated in the second experiment where gym attendance decreased among participants with regular baseline attendance (Charness & Gneezy, 2011).

Pope & Harvey (2015) conducted a RCT of first year college students to examine the impact of FIs on PA specific to SDT domains of extrinsic and intrinsic motivation over 24-weeks. Participants were randomized into control (n=39), continued- (n=39) and discontinued-incentive (n=39) conditions. In the continued-incentive condition, participants received FIs during fall and spring semesters whereas participants in the discontinued-incentive condition only received FIs in the fall. PA was measured by attendance at the campus fitness center and duration needed to exceed 30 minutes to count towards the weekly attendance goal. Intrinsic and extrinsic motivation was assessed using the Exercise Motivation Inventory-2 (EMI-2; Markland & Ingledew, 1997). During the fall semester, the control condition met 13% attendance goals, whereas the continued-incentive and discontinued-incentive met 62% and 64% of attendance goals, respectively. The difference between the control condition and FI conditions was significant χ^2 (1, n = 117) = 37.66, $p < 0.001$. In the spring semester, the control and discontinued-incentive condition met 3% of attendance goals, whereas the continued-incentive condition met 39% of attendance goals. The difference between the continued-incentive condition and the discontinued-incentive and control conditions was also significant χ^2 (2, n = 113) = 21.07, $p < 0.001$. Notably, there were no significant effects of condition on intrinsic domains (enjoyment and revitalization) and the extrinsic domain of appearance using the EMI-2. These results contradict the undermining effect and coincide with findings that FIs do not undermine intrinsic motivation for health behaviours when baseline participation and interest are low (Promberger & Marteau, 2013). However, findings must be interpreted with caution in the context of maintained PA after FI removal given the significant decline in attendance rates among participants in the discontinued-incentive condition (Pope & Harvey, 2015).

In a 8-week optimization trial, Kramer et al. (2020), evaluated intervention components of the Assistant to Lift your Level of activity (Ally) app that included FIs for meeting daily step goals, weekly planning, and daily self-monitoring prompts (Filler et al., 2015). The effects of FIs on intrinsic motivation were also explored. Insurees (n = 274) of a health insurance company in Switzerland were randomized into two FI conditions (cash and charity) and a control group at baseline. For the study duration, participants were randomized weekly into different planning conditions (action planning, coping planning, and no planning) and daily to receive or not receive a self-monitoring prompt. The primary outcome was achievement of personalized daily step goals and self-determined motivation was measured using the Behavioral Regulation in Exercise Questionnaire-2 (BREQ-2; Markland & Tobin, 2004). Results indicated that daily cash FIs significantly increased step goal achievement by 8.1% (95% CI 2.1 – 14.1; $p \leq 0.05$) during the 6-

week intervention period. Cash and charity FIs had no effect on post-intervention levels of intrinsic motivation, despite high degrees of baseline PA and intrinsic motivation. Although post-intervention PA was not assessed, the results suggest that FIs can increase daily step count without undermining intrinsic motivation even if baseline levels of intrinsic motivation and behaviour are high (Kramer et al., 2020).

J.M. Murray et al. (2020) performed a mediation analysis on the incentive-based PAL scheme cluster RCT to examine the short- (< 6 months) and long-term (≥ 6 months) mediation effect of psychosocial variables (e.g., intrinsic motivation) on PA. Participants (n = 853) were randomized into intervention (n = 457) and wait-list control (n = 396) conditions. Physical activity was assessed using pedometers at baseline, 6, and 12-months. Hypothesized short-term mediators were measured at baseline and 4-weeks while hypothesized long-term mediators were evaluated at baseline and 6-months. Results indicated a significant decrease in steps per day at 6-months in the intervention versus control group (adjusted MD: $b = -336$; $p = 0.02$) that was partially reduced by positive indirect effects through 6-month integrated regulation (between-group daily step MD attributable to mediator, adjusted for baseline: $ab = 94.7$ steps/d, 95% CI 18.7 – 204.4; $p < 0.05$), intrinsic motivation ($ab = 115.0$ steps/d, 95% CI 3.09 – 154.5; $p < 0.05$), and habit ($ab = 198.7$ steps/d, 95% CI 84.3 – 369.9; $p < 0.05$). There were no between-group differences in daily steps at 12-months, but positive indirect effects through 6-month integrated regulation ($ab = 128.0$ steps/d, 95% CI 27.3 – 313.2; $p < 0.05$), planning ($ab = 115.0$ steps/d, 95% CI 3.71 – 285.5; $p < 0.05$), and habit ($ab = 153.3$ steps/d, 95% CI 39.3 – 333.1; $p < 0.05$). While the overall intervention effects were negative, this was not explained by intrinsic motivation as predicted by the undermining effect when providing FIs. Furthermore, increased forms of internalized motivation (i.e., integrated regulation, intrinsic motivation) mitigated the negative effect and were associated with increased PA at 6 and 12-months (J. M. Murray et al., 2020).

Chapter 3

3 **Methodology**

3.1 ***Setting***

Carrot Rewards was a free mHealth app developed by *Carrot Insights Inc.* as part of a public-private partnership with the Public Health Agency of Canada and provincial/territorial Ministries of Health that incentivized Canadians for engaging in healthy behaviours, such as walking or completing educational health quizzes (Public Health Agency of Canada, 2015). The app was available for download to residents in BC, NL and ON in the Apple iTunes and Google Play app stores in March 2016, June 2016, and February 2017, respectively (Shankar, 2019). *Carrot Rewards* went out of business in 2019, at which time more than 1.5 million Canadians had downloaded the app (Marotta, 2019). In brief, higher than anticipated app engagement on such a large scale proved too costly for *Carrot Rewards*' government partners to fund (Rondina et al., 2020).

3.2 ***Program Description***

Carrot Rewards incentivized daily step count goal achievements with loyalty reward points worth $0.04 CAD per day (i.e., redeemable for consumer goods like movies or gas). Individualized daily step count goals were set by adding 500 to 1000 steps to users' 30-day daily step count median. After four weeks of earning rewards for meeting personalized daily step count goals, users could earn $0.40 CAD bonuses for completing longer "Step Up Challenges". Users could complete "Step Up Challenges" by reaching their daily step count goal ≥10 non-consecutive times over a 14-day period (Mitchell et al., 2018). In March 2018, “Step Together Challenges” were introduced as well which allowed users to pursue small team-based goals with a friend. Users participating in “Step Together Challenges” could also earn a \$0.40 CAD bonus for collaboratively reaching ≥10 individual daily step goals in a 7-day period with a friend (see Figure 1; Pearson et al., 2020). Finally, users could earn FIs by completing one to two short educational health quizzes per week about healthy living practices (e.g., physical activity, healthy eating) and self-regulatory healthy behaviour skills (e.g., goal setting, barrier identification; Mitchell et al., 2017). On December 08, 2018, the Government of Ontario ceased funding *Carrot Rewards* (Ng, 2018) in large part because of the cost of the intervention (approximately \$15 CAD per user per year; Rondina et al., 2021). Other revenue sources (e.g., private investors) allowed *Carrot Insights Inc.* to continue to offer the app to Ontarians free-of-charge (Marotta, 2019), but rewards for individual-level daily step goal achievements were discontinued as these drove intervention costs more than any other earning opportunity (i.e., about 80% of FIs earned were from daily step goal achievements; Rondina et al., 2020). Users in ON were informed five days prior to the withdrawal of rewards for individual-level daily step goal achievements (i.e., December 03, 2018) with an email from *Carrot Rewards* (Ng, 2018). At the same time, FIs for daily step count goal achievements persisted in BC and NL presenting a unique research opportunity to explore the impact of FI withdrawal on PA in a real-world public health setting.

Figure 1. Carrot Rewards “Steps” walking program interface.

Note. From “Adding team-based financial incentives to the Carrot Rewards physical activity app increases daily step count on a population scale: a 24-week matched case control study,” by Pearson et al., 2020, International Journal of Behavioral Nutrition and Physical Activity, 17(1), p. 139 (https://doi.org/10.1186/s12966-020-01043-1). Copyright © [2020] by Pearson et al.

3.3 ***Study Design***

To examine this ‘naturally occurring experiment’, a 25-week QE was adopted using a retrospective pre-post design with non-equivalent control groups (i.e., intervention group

= 438,731 *Carrot Rewards* users in ON vs. control groups = *Carrot Rewards* users in BC [n=124,101] and NL [n=21,928]). To conduct public health research that offers greater opportunity for adaptation and iterative refinement of protocols and intervention delivery, our design was selected to align with the maintenance dimension of the RE-AIM framework. RE-AIM is a public health program evaluation framework that addresses five dimensions (i.e., reach, effectiveness, adoption, implementation, and maintenance) of individual- and setting-level outcomes critical to intervention impact and sustainability (Glasgow et al., 1999). At the individual-level, maintenance has been defined as “the long-term effects of a program on outcomes after 6 or more months after intervention contact” (Kwan et al., 2019). Our 25-week QE design assessed the effects of FIs on PA in *Carrot Rewards* users with a maximum app exposure of 30-, 27-, and 19-months in BC, NL, and ON, respectively. Performing RCTs in fast-paced digital health settings can be challenging, but the mHealth field has benefitted from QE designs that attempt to determine causality for outcomes of intervention effectiveness (Handley et al., 2018). When RCTs are infeasible, QEs can be exploited to evaluate causal effects (Kim & Steiner, 2016) that may contribute to an understanding of the contextual (e.g., user engagement) and program (e.g., FI removal) factors that impact PA and ultimately influence intervention effectiveness (Brower et al., 2020).

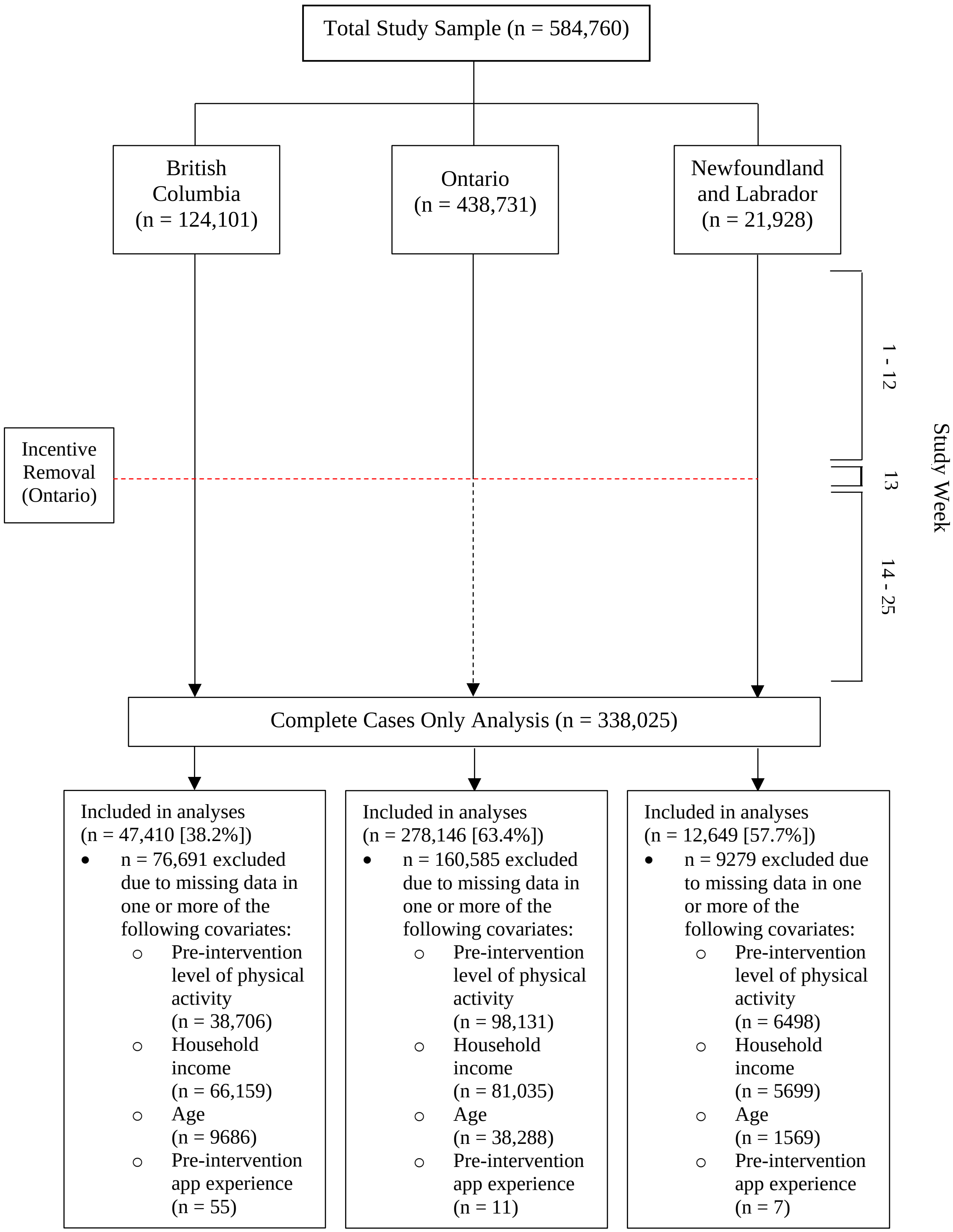
Pre-post designs with non-equivalent control groups examine the effect of an intervention by concurrently comparing pre- and post-intervention period differences between intervention and control groups (Handley et al., 2018). In our study, the ‘intervention’ occurred at the end of Study Week 13 when FIs for individual-level daily step goal achievements were *withdrawn* in ON (December 8, 2018; see Fig. 2). The intervention period was defined as the start of Study Week 13 to the end of Study Week 17 (December 2, 2018 to January 5, 2019) to minimize potential threats to internal validity from the anticipation effect (Waddington et al., 2017) and history bias (Naci & Soumerai, 2016). The anticipation effect refers to potential changes in behaviour and outcomes that result from knowledge of a future intervention prior to implementation (Waddington et al., 2017). *Carrot Rewards* users were informed of FI removal on December 3, 2018 (Ng, 2018) which may have differentially influenced daily step count in ON relative to BC and NL prior to the intervention. To control for bias associated with the anticipation effect as a result of prior knowledge of FI removal, each day of Study Week 13 leading up to the intervention (i.e., December 2 to 7, 2018) was excluded from the pre-intervention period and included in the intervention period. History bias, on the other hand, refers to co-occurring events before, during, or after the intervention period that are unrelated to an intervention but effect outcomes (Naci & Soumerai, 2016). During the Canadian winter holiday season, PA has been shown to sharply and predictably decline across Canada (McGavock et al., 2019) which could have extraneously influenced step count after the intervention. To minimize history bias associated with the Canadian winter holiday season, the start of Study Week 14 to the end of Study Week 17 (December 9, 2018, to January 5, 2019) was excluded from the post-intervention period (and included in the intervention period). Furthermore, 25-weeks of daily step count measurements were included in our analyses to approximate an interrupted time series design and enable examination of potential threats from history bias (Handley et al., 2018). The intervention period was accounted for in analyses by specifying a separate intervention period level for each of Study Weeks 13 to 17 (see shaded area in Fig. 3). Therefore, step count data that may have been influenced by the anticipation effect and history bias were excluded from analyses of the pre- and post-intervention period differences between provinces used to assess the intervention effect.

The ‘pre-intervention’ period was defined as the 12 weeks preceding FI withdrawal (Study Weeks 1 to 12; September 9 to December 1, 2018). The ‘post-intervention’ period was defined as the final eight weeks of the study (Study Weeks 18 to 25; January 6 to March 2, 2018). During the post-intervention period, ON users could continue to earn rewards for completing harder-to-achieve “Step Together Challenges” as well as educational health surveys. It is estimated that users in ON earned approximately $1.56 CAD during the post-intervention period compared to $4.46 CAD in BC and NL.

To further increase the internal validity of study results, selection bias was also considered in the design of this retrospective pre-post study with non-equivalent control groups (Handley et al., 2018). Selection bias concerns meaningful differences between intervention and non-equivalent control group sites (i.e., sociodemographic differences between ON and BC/NL) that can impact the outcome of interest (i.e., post-intervention step count) and bias results (Nunan et al., 2017). Selection bias was addressed, in part, by balancing measures of pre-intervention period behaviour (i.e., app experience, app engagement, and level of PA) and baseline participant characteristics (i.e. age, gender, household income, loyalty rewards program, and baseline step count) between provinces in analyses (Brazauskas & Logan, 2016; Handley et al., 2018).

Ethical approval for this study was provided by the Western University Human Research Ethics Board (#114790; see Appendix A).

Figure 2. Study flow chart.



3.4 ***Outcome Measure***

The outcome measure was weekly mean daily step count measured by built-in smartphone accelerometers (Study Weeks 1-25). Validation studies have shown that the step counting feature in iPhone and Android smartphones are accurate when compared to gold standards of measurement such as manual step counting and research-grade accelerometers in laboratory conditions (Duncan et al., 2017; Hekler et al., 2015). For instance, in a laboratory condition Duncan et al. (2017) found that at speeds above 5 kilometers per hour (km/h) the mean bias of the iPhone step counting feature when compared to manually counted steps was within acceptable (< ± 5%) levels required of research-grade pedometers (Tudor-locke et al., 2006; Vincent & Sidman, 2003; Welk et al., 2000). In field conditions, however, iPhones have been shown to significantly underestimate steps compared to research-grade accelerometers by about 20% (or 1340 steps/day) on average and fail to meet acceptable levels of mean bias (± 10%) established in previous free-living studies (Barreira et al., 2013; Schneider et al., 2004). The inconsistency between laboratory- and field-based studies is largely attributed to participant behaviour, such as carrying method (i.e., location on body or use of a bag) and "wear time" (i.e., daily carrying adherence; Duncan et al., 2017). Caution should be exercised when using smartphones instead of research-grade pedometers/accelerometers to measure PA for research purposes. However, if adherence (i.e., "wear time") can be increased with, for example, multi-component mHealth apps that include educational content, daily/weekly PA goals, biofeedback, FIs, etc., then it has been suggested that smartphones may be suitable for PA evaluations (Duncan et al., 2017).

3.5 ***Covariates***

The FI for PA literature informed covariate selection *a priori* given their potential role in moderating the impact of FI removal on PA, including: (1) pre-intervention PA level (Mitchell, Orstad, et al., 2020), (2) pre-intervention app engagement (Mitchell et al., 2018; Mitchell, Lau, et al., 2020; J. M. Murray et al., 2019; Voils et al., 2012), (3) app experience (Prochaska & Velicer, 1997; Voils et al., 2012), and (4) socio-demographics and other participant characteristics (Mitchell, Orstad, et al., 2020; Wurst et al., 2020). First, pre-intervention PA was calculated using the average of weekly mean daily step count during Study Weeks 1 to 12 where at least one measure of weekly mean daily step count was required. Participants were then stratified into PA tertiles using thresholds defined by Tudor-Locke et al. (2013): Sedentary users = < 5000 steps/d, low active users = 5000 to 7499 steps/d, and physically active users = ≥ 7500 steps/d. Second, pre-intervention app engagement was determined by counting the number of weeks users opened the app at least once during Study Weeks 1 to 12 (mHealth app engagement subdimension of ‘frequency’; Voils et al., 2012). Participants were stratified into pre-intervention app engagement tertiles (low engagers = 0 to 4 weeks, medium engagers = 5 to 8 weeks, and high engagers = > 8 weeks). Next, app experience was determined by counting the number of months users had engaged with the *Carrot Rewards* app prior to Study Week 12 (mHealth app engagement subdimension of ‘amount’; Voils et al., 2012). Participants were stratified into app experience tertiles (low experience = < 6 months, medium experience = 6 to 12 months, and high experience = > 12 months). Notably, the maintenance stage of change defined by the transtheoretical model of behaviour change begins at six months (Prochaska & Velicer, 1997). Age and gender were included as covariates as well to determine whether the purported influence of either sociodemographic characteristic on PA *during* FI provision (Mitchell, Orstad, et al., 2020; Wurst et al., 2020) extended to PA *after* FI removal in this study. Participants between the ages of 26 and 85 years were categorized into six 10-year cohorts. Adolescents and younger adults were categorized by cohorts spanning the ages of 13 – 17, and 18 to 26 years, respectively. Participants older than 85 years were categorized into a single cohort. In the present study, participants were categorized as either female, male, or other (i.e., identified gender not male or female). We are examining the role of gender in this study – that is, the socialized gendered identification of participants. For this relatively short-term project, the gendered effect is what we would really like to establish and sex at birth or the view of sex as a biological difference will not impact the way a person will engage with the app or their subsequent behaviours (i.e., PA or daily steps taken; Natural Sciences and Engineering Research Council of Canada, 2017). In addition to age and gender, loyalty rewards program, household income and baseline daily step count were included as covariates. Participants could earn FIs from their loyalty rewards program of choice (i.e., RBC Rewards®, SCENE® Points, Aeroplan® Miles, Drop Points, Petro-Points™, and More Rewards®). Household income was self-reported and participant baseline daily step count was measured upon initial app download using the mean daily step count assessed during a 7-day baseline period.

3.6 ***Statistical Analyses***

3.6.1 ***Sample Characteristics.*** A Chi-square test of independence was conducted on categorial baseline characteristics to determine if there were any discrepancies in gender, household income and loyalty rewards program between provinces (García-Pérez &

Núñez-Antón, 2003). To reduce the likelihood of Type 1 errors which was increased by conducting simultaneous Chi-squared analyses for each baseline characteristic between provinces, estimated *p*-values were compared against Bonferroni corrected *p*-values. The Bonferroni corrected *p*-values were equal to p / n, where p equaled the level of significance and n equaled the total number of comparisons for each categorical baseline characteristic between provinces (Shaffer, 1995). Level of significance was set to p = 0.05 and divided by n, equal to the product of provinces (n = 3) and subgroups of gender (n = 3), household income (n = 10), and loyalty rewards program (n = 6). Accordingly, the Bonferroni corrected *p*-values for gender, household income, and loyalty rewards program equaled 0.0055 (0.05/9), 0.0017 (0.05/30), and 0.0028 (0.05/18), respectively. Continuous baseline characteristics (i.e., age and baseline daily step count) and pre-intervention period behaviours (i.e., PA level, app engagement, app experience) were analyzed using the Independent-Samples Kruskal-Wallis test. The Independent-Samples Kruskal-Wallis test is a non-parametric analysis of quantitative outcomes in three or more groups and was used because sample data were not normally distributed (Kruskal & Wallis, 1952). Tests of normality indicated that age was moderately skewed (0.795) and platykurtic (0.152) while baseline step count was highly skewed (1.282) and platykurtic (2.836). Pre-intervention app engagement was highly skewed (-1.210) and platykurtic (-0.003) whereas level of PA was highly skewed (1.289) but leptokurtic (3.473). Although the skew of pre-intervention app experience was approximately symmetric (0.353) its distribution was platykurtic (-0.485; Balanda & Macgillivray, 1988; Bulmer, 1967).

3.6.2 ***Primary and Secondary Analyses.*** The primary study objective was to examine the impact of daily FI removal on weekly mean daily step counts in ON compared to BC and

NL where FI availability did not change. To address the primary study objective, the two-way interaction between study week and intervention period on weekly mean daily step count was examined with a simple linear regression model for each province. All covariates were included in the models as additive effects (American Psychological Association, n.d.; Coz, 1984) to minimize selection bias by balancing covariates between provinces (Brazauskas & Logan, 2016; Handley et al., 2018). Physical activity level, app engagement, app experience, and age were included in the models as continuous covariates. The simple linear regression model used for the primary analysis is presented as Equation 1 in Appendix B.

The secondary study objectives were to explore whether covariates (e.g., PA and mHealth app engagement levels, age) influenced the impact of daily FI removal on weekly mean daily step counts in ON compared to BC and NL. To address the secondary study objectives, the three-way interaction between covariate level, Study Week, and intervention period on weekly mean daily step count was examined with simple linear regression models for each province. Separate models were used to analyze the three-way interaction for each covariate of interest (i.e., pre-intervention behaviours: PA level, app engagement, and app experience tertiles; baseline sociodemographic characteristics: age cohorts and gender categories). When a covariate was not analyzed for a three-way interaction, it was included in each model as an additive effect along with baseline step count, household income, and loyalty rewards program (American Psychological Association, n.d.; Coz, 1984) to minimize selection bias by balancing covariates between provinces (Brazauskas & Logan, 2016; Handley et al., 2018). When included in models as an additive effect, the continuous values of PA level, app engagement, app experience and age were used. The simple linear regression model used for the secondary analyses is presented as Equation 2 in Appendix B.

Simple linear regression was performed with a robust sandwich estimator of covariance to account for variance and correlation within each user over study weeks. The robust sandwich estimator of covariance specifies a heteroskedastic covariance model that does not assume constant variance and uncorrelated measurements between time points which improves the accuracy of SEs of estimated coefficients and CIs for repeated measures data (Fitzmaurice et al., 2011). The acute impact of the intervention (i.e., change in PA level) on weekly mean daily step count was assessed by calculating the difference of the pre- (Study Week 12;) and post-intervention (Study Week 21) intercepts (γ_{12} and γ_{21}, respectively) within and between provinces. γ_{12} was specified to correspond with the last measurement of the pre-intervention period prior to intervention implementation. To allow time for the intervention to take effect, γ_{21} was defined as the midpoint of the post-intervention period. To estimate γ_{12} and γ_{21}, the mean of continuous and the proportion of categorical covariates among users at Study Weeks 12 and 21 (i.e., $t = 12$ and $t = 21$, respectively) were inputted into models for the primary and secondary analyses. The impact of FI removal over time (i.e., rate of change) on weekly mean daily step count was measured by calculating the difference in slope of the pre- (i.e., β_1) and post-intervention ($\beta_1 + \beta_{13}$) periods within and between provinces. Comparisons within province were conducted to assess the direct effect of the intervention in ON relative to BC and NL. Comparisons between provinces were performed to evaluate the size of the intervention effect in ON relative to BC/NL and to make a direct comparison between BC and NL where the intervention did not occur. Estimated slope and intercept of weekly mean daily step count were independent for each province and hypothesized to be normally distributed given the large sample sizes. Although tests of normality indicated that covariates (i.e., pre-intervention app engagement) and certain participant characteristics (e.g., baseline step count) were not normally distributed, the central limit theorem justifies the use of parametric tests when analyzing groups with large sample sizes (i.e., n > 40) even if the data is non-normal (Elliott & Woodward, 2007). Therefore, the estimated change in slope and intercept of weekly mean daily step count between the pre- and post-intervention period were compared using estimated SEs from each province to calculate the unpooled variance for the estimate of their differences. The primary and secondary analyses were completed using the *lm* base function and *vcovCR* function from the package *'clubSandwich'* in *RStudio* version 4.0.5 (RStudio, Boston, MA, USA).

3.6.3 ***Complete Case Analysis.*** The total study sample (n = 584,760) was comprised of users in ON, BC, and NL with and without missing data for variables required for analyses. As has been suggested previously, a complete case (CC) analysis and multiple imputation (MI) were used to handle missing data (Sterne et al., 2009). Risk of bias associated with mechanisms of missingness was ascertained by comparing results of the primary analysis from CC and MI (Sterne et al., 2009). First, a CC analysis was used to select an analytic sample from the total study sample. Complete case analysis is the default option for missing data analysis in statistical software packages (White & Carlin, 2010) and is less computationally intensive than MI (Sterne et al., 2009). The CC sample was selected under the assumption that the mechanism of missingness was completely at random (MCAR). Data that is MCAR is not related to any observed and unobserved variables (Little & Rubin, 2014). Contingent upon the data being MCAR, the CC sample was equivalent to a random sample of the total study sample which would not bias results of the analyses (Little & Rubin, 2014). Users in the total study sample were excluded from the CC sample if they did not have complete measures of each covariate. By virtue of the calculation for pre-intervention PA level, users in the CC sample were required to have at least one measure of mean weekly daily step count during the pre-intervention period (Study Weeks 1 – 12).

A Chi-squared test of independence was performed to determine if there were any discrepancies in categorial baseline characteristics between users who were included and excluded from the analytic sample. Chi-squared statistics for each baseline characteristic between users who were included/excluded from the CC sample were also compared to Bonferroni corrected *p*-values to reduce the likelihood of Type 1 errors (Shaffer, 1995). Level of significance was set to $p = 0.05$ and divided by n, equal to the product of users who were included/excluded from the CC sample (n = 2) and subgroups of gender (n = 3), household income (n = 11; additional subgroup for non-applicable), and loyalty rewards program (n = 6). Accordingly, the Bonferroni corrected *p*-values for gender, household income, and loyalty rewards program equaled 0.0083 (0.05/6), 0.0023 (0.05/22), and 0.0042 (0.05/18), respectively.

Continuous baseline characteristics (i.e., age and baseline step count) and pre-intervention period behaviours (i.e., PA level, app engagement, app experience) were analyzed using the Independent-Samples Kruskal-Wallis test. The *complete.cases* function in *RStudio* version 4.0.5 (RStudio, Boston, MA, USA) was used to select the CC sample which consisted of 338,025 users in ON, BC, and NL. Forty-two percent (n = 246,735) of users in the total study sample were excluded from the CC sample for meeting exclusion criteria, with more users in BC being excluded (61.8%) than NL and ON (42.3% and 36.6%, respectively; see Fig. 2). The aggregate effect of missing data for several variables led to an exclusion of a significant proportion of the total study sample from the CC analysis. Even if the MCAR assumption was correct, the reduced size of the analytic sample decreased study power and precision. Furthermore, if data was not MCAR the CC sample would not be representative of the total study sample which could produce biased and imprecise results (Sterne et al., 2009).

3.6.4 ***Total Sample (Sensitivity) Analysis***. Acknowledging some of the limitations of the CC analysis (i.e., violation of the MCAR assumption and reduced sample size), the second statistical method used to address problems from missing data was multiple imputation (MI). The sensitivity of the CC analysis to the MCAR assumption was examined by comparing results of the primary analysis between the CC sample and the total study sample, where MI was used to impute missing data (Fig. 2). MI requires data to be missing at random (MAR), where missingness is conditional on observed variables and independent of unobserved variables. Using five iterations, five imputed datasets were created by imputing the missing data of effected variables through sampling from the predicted distribution of observed data (Little & Rubin, 2014). As a repeated measure and continuous variable, weekly mean daily step count was imputed using a heteroscedastic linear two-level model by a Gibbs sampler (Kasim & Raudenbush, 1998; van Buuren & Groothuis-Oudshoorn, 2011). Continuous covariates and sociodemographic characteristics were imputed using predictive mean matching. Categorical sociodemographic characteristics were imputed using polytomous logistic regression (van Buuren & Groothuis-Oudshoorn, 2011). The simple linear regression model using a robust sandwich estimator of covariance for the primary analysis was fit to each imputed dataset to generate SEs and unpooled variances for calculations of estimated intercept and slope differences within and between provinces. Results of the simple linear regression model with a robust sandwich estimator of covariance using the multiply imputed data and the CC sample were then compared. MI and analysis of the multiply imputed data were carried out using the *mice* and *bucky* packages, respectively, in *RStudio* version 4.0.5 (RStudio, Boston, MA, USA).

Chapter 4

4 Results

4.1 *Sample Characteristics*

The CC sample consisted of 338,025 participants (57.8% of the total sample) from ON (n = 278,146), BC (n = 47,410), and NL (n = 12,469). Significant provincial differences in age, gender, household income, loyalty rewards program and baseline daily step count were noted (Table 1). Pre-intervention app experience, engagement, and level of PA are shown in Table 2. Participants from BC were more engaged in the pre-intervention period than those from ON and NL. Participants from ON also had less app experience than the other provinces owing to the later app launch in ON. Regarding PA levels, significant provincial differences were noted in the pre-intervention period with participants from NL accumulating fewer steps per day (*M* = 5863 steps/d, *SD* = 3124) compared to those from ON (*M* = 6431 steps/d, *SD* = 3058) and BC (*M* = 6712 steps/d, *SD* = 3181). Characteristics of the total study sample (n=584,760) including users with and without missing data are also presented in Appendix C. Notably, mean age (32.15 years [yrs] vs. 35.40 yrs) and pre-intervention app engagement (8.02 weeks [wks] vs. 9.62 wks) were lower in the total compared to the CC sample.

Table 1. Baseline characteristics, complete cases sample.

Variable	Ontario (n = 278,146)	British Columbia (n = 47,410)	Newfoundland and Labrador (n = 12,469)
Age (years; mean ± SD)[a]	33.92 ± 12.72*	36.51 ± 13.33*	35.78 ± 12.77*
Gender[b]			
Female	179,744 (64.6%)*	31,684 (66.8%)*	8960 (71.9%)*
Male	94,365 (33.9%)*	14,784 (31.2%)*	3398 (27.2%)*
Other	4037 (1.5%)*	942 (2.0%)*	111 (0.9%)*
Household Income (CAD/year)[b]			
< 20,000	25,896 (9.3%)*	3868 (8.2%)*	1083 (8.7%)
20,000 > 40,000	34,575 (12.4%)	5919 (12.5%)	1474 (11.8%)
40,000 > 60,000	42,182 (15.2%)	7777 (16.4%)*	1725 (13.8%)
60,000 > 80,000	35,961 (12.9%)	6375 (13.5%)	1494 (12.0%)
80,000 > 100,000	29,271 (10.5%)	5039 (10.6%)	1374 (11.0%)
100,000 > 150,000	31,579 (11.4%)*	5741 (12.1%)	1705 (13.7%)*
≥ 150,000	23,030 (8.3%)*	3321 (7.0%)*	1120 (9.0%)
Didn't Complete Survey	3220 (1.2%)*	337 (0.7%)*	121 (1.0%)
Don't Know	11,172 (4.0%)*	1470 (3.1%)*	366 (2.9%)
Rather Not Say	41,260 (14.8%)*	7563 (15.9%)	2007 (16.1%)
Loyalty Rewards Program[b]			
Aeroplan® Miles	41,084 (14.8%)*	8595 (18.1%)*	3320 (26.6%)*
Drop Points	12,605 (4.5%)*	1615 (3.4%)*	598 (4.8%)
More Rewards®	401 (0.1%)*	4856 (10.3%)*	22 (0.2%)*
Petro-Points™	30,244 (10.9%)*	2989 (6.3%)*	87 (0.7%)*
RBC Rewards®	5724 (2.1%)	1040 (2.2%)	212 (1.7%)
SCENE Points®	188,088 (67.6%)*	28,315 (59.7%)*	8230 (66.0%)
Baseline step count[c] *(steps/day; mean ± SD)*[a]	5751 ± 3714*	5883 ± 3513*	5307 ± 3485*

Note. *SD* = standard deviation. *CAD* = Canadian dollars.

[a] = Independent-Samples Kruskal-Wallis Test; [b] = Chi-squared test of independence;

[c] = mean daily step count over 14-days prior to Study Week 1.

* = $p < .05$.

Table 2. Pre-intervention period behaviours, complete cases sample.

Variable	Ontario (n = 278,146)	British Columbia (n = 47,410)	Newfoundland and Labrador (n = 12,469)
App Engagement[a] *(weeks; mean ± SD)*[b]	$9.37 \pm 3.74_a$	$10.11 \pm 3.17^*$	$9.38 \pm 3.71_a$
App Experience[c] *(months; mean ± SD)*[b]	$12.75 \pm 5.57^*$	$17.54 \pm 8.45_b$	$17.93 \pm 8.93_b$
Level of Physical Activity (weekly mean daily step count; mean ± SD)[b]	$6431 \pm 3058^*$	$6712 \pm 3181^*$	$5863 \pm 3124^*$

Note. *SD* = standard deviation. Means sharing a common subscript were not significantly different at $p < .05$ according to the Independent-Samples Kruskal-Wallis test.

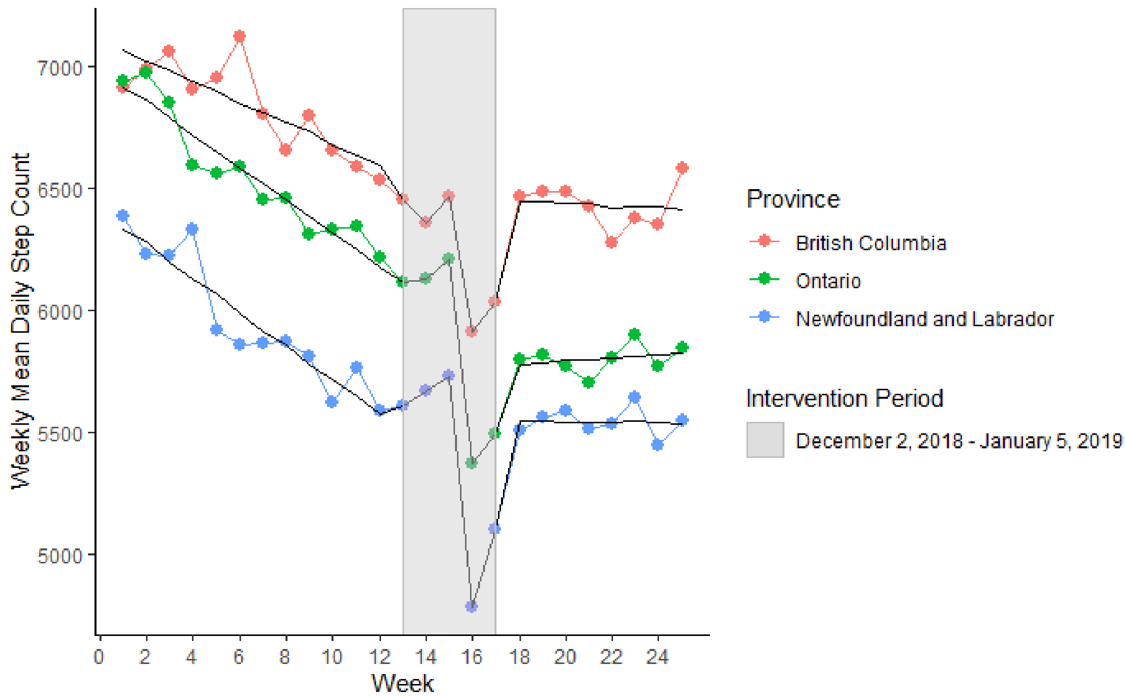
[a] = weeks the app was opened at least once during the pre-intervention period (Study Weeks 1 – 12). [b] = Independent-Samples Kruskal-Wallis Test; [c] = months since “Steps” program enabled prior to Week 12.

* = $p < .05$.

4.2 ***Primary Analyses***

Observed changes in weekly mean daily step count fit with a simple linear regression model of the two-way interaction between week and intervention period for the CC sample of each province is illustrated in Figure 3. Estimates of pre- and post-intervention weekly mean daily step count intercepts (γ_{12}, and γ_{21}, respectively) and slopes (β_1, and $\beta_1 + \beta_{13}$, respectively) are provided in Tables 3 and 4. Notably, estimated intercept values dropped from pre- to post-intervention in all three provinces with the most pronounced decrease noted in ON (ON: $\gamma_{12} - \gamma_{21} = -367$ steps/d, $p < .01$; BC: $\gamma_{12} - \gamma_{21} = -169$ steps/d, $p < .01$; NL: $\gamma_{12} - \gamma_{21} = -93$ steps/d, $p < .01$; Table 3). In addition, the pre- to post-intervention intercept difference was greatest when comparing ON to BC and NL (198.4 and 274.1 steps/d, respectively; Table 4). Regarding weekly mean daily step count slopes, significant differences in post-intervention slope were observed in ON ($\beta_1 + \beta_{13} = 8.318$ steps/wk, $SE = 0.823$, 95% CI [6.71, 9.93], $p < .01$) and BC ($\beta_1 + \beta_{13} = -4.364$ steps/wk, $SE = -1.926$ [-8.14, -0.59], $p < .05$), but not NL ($\beta_1 + \beta_{13} = -2.174$ steps/wk, $SE = 3.519$ [-9.07, 4.72]); Table 3). Between provinces analyses show that the *positive* estimated post-intervention slope in ON was significantly different from BC ($\beta_1 + \beta_{13} = 12.68$ steps/wk, $SE = 2.094$ [8.58, 16.80], $p < .01$) and NL ($\beta_1 + \beta_{13} = 10.49$ steps/wk, $SE = 3.614$ [4.41, 17.57], $p < .01$; Table 4), though the rate of change was modest in terms of steps/d (difference of 1.81 and 1.50, respectively).

Figure 3. Provincial weekly mean daily step count by week and intervention period, complete cases sample.



Note. Observed averages (points) and averages of predictions from the simple linear regression model fit with the two-way interaction between week and intervention period (black line) of weekly mean daily step count by week in each province. The intervention period (Study Week 13 to 17) from December 2, 2018, to January 5, 2019, was accounted for in the regression model by specifying a separate intervention period level for each of the weeks 13 to 17. The pre- and post-intervention periods included Weeks 1 -12 and Weeks 18 – 25, respectively. The intervention occurred during Week 13 (December 8, 2018).

Table 3. Estimated weekly mean daily step count intercepts and slopes (within provinces), complete cases sample.

Parameter	Ontario			British Columbia			Newfoundland and Labrador		
	$\hat{B}$	*SE*	95% CI	$\hat{B}$	*SE*	95% CI	$\hat{B}$	*SE*	95% CI
Intercept									
Pre-[a]	6153**	3.045	[6147, 6159]	6535**	7.265	[6520, 6549]	5574**	14.02	[5547, 5602]
Post-[b]	5786**	4.344	[5777, 5794]	6366**	9.616	[6347, 6385]	5481**	19.75	[5442, 5520]
Slope									
Pre-[c]	-62.70**	-0.470	[-63.62, -61.78]	-41.90**	1.196	[-44.24, -39.55]	-62.10**	2.208	[-66.43, -57.77]
Post-[d]	8.318**	0.823	[6.71, 9.93]	-4.364*	-1.926	[-8.14, -0.59]	-2.174	3.519	[-9.07, 4.72]

Note. $\hat{B}$ = unstandardized regression coefficient; *SE* = robust standard error; CI = confidence interval.

[a] = study week 12; [b] = study week 21; [c] = study weeks 1 - 12; [d] = study weeks 18 - 25.

* = $p < .05$; ** = $p < .01$.

Table 4. Estimated weekly mean daily step count intercepts and slopes (between provinces), complete cases sample.

Parameter	Ontario and British Columbia			Ontario and Newfoundland and Labrador			British Columbia and Newfoundland and Labrador		
	$\hat{B}$	*SE*	95% CI	$\hat{B}$	*SE*	95% CI	$\hat{B}$	*SE*	95% CI
Intercept									
Pre-[a]	-381.3**	7.877	[-396.7, -365.8]	579.1**	14.35	[551.0, 607.2]	960.3**	15.79	[929.4, 991.3]
Post-[b]	-579.7**	10.55	[-600.4, -559.0]	305.0**	20.22	[265.4, 344.7]	884.7**	21.97	[841.7, 927.8]
Slope									
Pre-[c]	-20.81**	1.285	[-23.33, -18.29]	-0.603	2.257	[-5.03, 3.82]	20.20**	2.511	[15.28, 25.13]
Post-[d]	12.68**	2.094	[8.58, 16.8]	10.49**	3.614	[3.41, 17.57]	-2.19	4.012	[-10.05, 5.67]

Note. $\hat{B}$ = unstandardized regression coefficient; *SE* = robust standard error; CI = confidence interval.

[a] = study week 12; [b] = study week 21; [c] = study weeks 1 - 12; [d] = study weeks 18 - 25.

* = $p < .05$; ** = $p < .01$.

4.3 ***Secondary Analyses***

Estimated pre- and post-intervention weekly mean daily step count intercepts and slopes by covariate level of the CC sample are shown in Tables 5 and 6 (for app engagement) as well as Appendices E to H (for pre-intervention PA, app experience, age, and gender covariates). Notably, the estimated intercept decrease from pre- to post-intervention was more pronounced amongst highly engaged and physically active users in ON (high engagement: $\gamma_{12} - \gamma_{21}$ = -328 steps/d, $p < .01$; low engagement: $\gamma_{12} - \gamma_{21}$ = -211 steps/d, $p < .01$; physically active: $\gamma_{12} - \gamma_{21}$ = -232 steps/d, $p < .01$; sedentary: $\gamma_{12} - \gamma_{21}$ = 107 steps/d, $p < .01$). Sedentary users were the only covariate level to exhibit an increase in estimated intercept from pre- to post-intervention in ON. As well, post-intervention estimated slope in ON was lower among more highly engaged and physically active users (high engagement: $\beta_{35} + \beta_{41}$ = 7.538 steps/wk, *SE* = 0.860 [5.85, 9.22], $p < .01$; low engagement: $\beta_{21} + \beta_{27}$ = 24.23 steps/wk, *SE* = 4.417 [15.58, 32.89], $p < .05$; physically active: $\beta_{35} + \beta_{41}$ = -10.98 steps/wk, *SE* = 1.957 [-14.82, -7.15], $p < .05$; sedentary: $\beta_{21} + \beta_{27}$ = 24.75 steps/wk, *SE* = 1.076 [22.64, 26.86], $p < .05$), though the rates of change were modest in terms of steps/d (1.08, 3.46, -1.57 and 3.54, respectively). Level of app experience, age, and gender did not appear to influence the estimated intercept decrease from pre- to post-intervention in ON (maximum between-level difference: high - low experience = 14 steps/d; [56 – 65] – [18 – 25] yrs of age = 53 steps/d; female – male = 10 steps/d; Appendices F to H). While the post-intervention estimated slope significantly increased for certain levels of app experience, age, and gender within ON, the highest rates of change (high experience: $\beta_{35} + \beta_{41}$ = 10.85 steps/wk, *SE* = 1.119 [8.65, 13.04], $p < .01$; 26 – 35 yrs of age: $\beta_{35} + \beta_{41}$ = 18.75 steps/wk, *SE* = 1.524 [15.77, 21.74], $p < .01$;

other identified gender: $\beta_{35} + \beta_{41}$ = 14.78 steps/wk, *SE* = 6.693 [1.66, 27.90], $p < .05$; Appendices F to H) were modest in terms of steps/d (1.55, 2.68 and 2.11, respectively).

Table 5. Estimated weekly mean daily step count intercepts and slopes by application engagement (within provinces), complete cases sample.

Parameter	Ontario			British Columbia			Newfoundland and Labrador		
	$\widehat{B}$	*SE*	95% CI	$\widehat{B}$	*SE*	95% CI	$\widehat{B}$	*SE*	95% CI
Intercept									
Pre-[a]									
Low[b]	6263**	11.84	[6239, 6286]	6636**	40.47	[6556, 6715]	5622**	52.40	[5519, 5725]
Medium[b]	6255**	10.35	[6235, 6275]	6629**	35.95	[6559, 6700]	5632**	45.83	[5543, 5722]
High[b]	6149**	4.117	[6141, 6157]	6526**	10.35	[6506, 6546]	5552**	18.11	[5517, 5588]
Post-[c]									
Low[b]	6052**	19.13	[6015, 6090]	6704**	52.28	[6601, 6806]	5742**	88.63	[5568, 5916]
Medium[b]	6036**	16.74	[6003, 6069]	6680**	46.48	[6589, 6771]	5719**	77.68	[5566, 5871]
High[b]	5821**	6.547	[5808, 5833]	6400**	13.69	[6373, 6427]	5520**	29.60	[5462, 5578]
Slope									
Pre-[d]									
Low[b]	-46.63**	1.608	[-49.78, -43.48]	-20.81**	5.482	[-31.56, -10.07]	-52.55**	7.971	[-68.17, -36.92]
Medium[b]	-60.44**	1.436	[-63.25, -57.62]	-39.25**	3.987	[-47.06, -31.43]	-46.17**	6.458	[-58.83, -33.52]
High[b]	-64.55**	0.516	[-65.56, -63.54]	-43.35**	1.279	[-45.86, -40.85]	-64.82**	2.419	[-69.56, -60.08]

Table 5 (continued).

Parameter	Ontario			British Columbia			Newfoundland and Labrador		
	$\hat{B}$	*SE*	95% CI	$\hat{B}$	*SE*	95% CI	$\hat{B}$	*SE*	95% CI
Slope									
Post-[e]									
Low[b]	24.23**	4.417	[15.58, 32.89]	6.018	11.44	[-16.41, 28.45]	8.503	18.16	[-27.09, 44.10]
Medium[b]	10.99**	3.471	[4.19, 17.80]	4.828	7.669	[-10.20, 19.86]	1.691	14.57	[-26.87, 30.25]
High[b]	7.538**	0.860	[5.85, 9.22]	-5.512*	2.018	[-9.47, -1.56]	-3.004	3.688	[-10.23, 4.23]

Note. $\hat{B}$ = unstandardized regression coefficient; *SE* = robust standard error; CI = confidence interval.

[a] = study week 12; [b] = $0 \leq 4$, $5 \leq 8$, and $9 \leq 12$ weeks the application was opened at least once pre-intervention for low, medium, and high engagement, respectively; [c] = study week 21; [d] = study weeks 1 - 12; [e] = study weeks 18 - 25.

* = $p < .05$; ** = $p < .01$.

Table 6. Estimated weekly mean daily step count intercepts and slopes by application engagement (between provinces), complete cases sample.

Parameter	**Ontario and British Columbia**			**Ontario and Newfoundland and Labrador**			**British Columbia and Newfoundland and Labrador**		
	$\hat{B}$	***SE***	**95% CI**	$\hat{B}$	***SE***	**95% CI**	$\hat{B}$	***SE***	**95% CI**
Intercept									
Pre-[a]									
Low[b]	-373.1**	42.16	[-455.8, -290.5]	640.4**	53.72	[535.1, 745.7]	1014**	66.20	[883.8, 1143]
Medium[b]	-374.5**	37.41	[-447.9, -301.2]	622.5**	46.98	[530.4, 714.6]	997.1**	58.25	[882.9, 1111]
High[b]	-377.3**	11.14	[-399.1, -355.5]	596.3**	18.57	[559.9, 632.7]	973.6**	20.86	[932.7, 1014]
Post-[c]									
Low[b]	-651.7**	55.67	[-760.8, -542.6]	310.5**	90.67	[132.7, 488.2]	962.2**	102.9	[760.5, 1164]
Medium[b]	-643.5**	49.40	[-740.3, -546.7]	317.7**	79.46	[161.9, 473.4]	961.2**	90.52	[783.8, 1139]
High[b]	-579.8**	15.18	[609.5, -550.0]	300.7**	30.32	[241.2, 360.1]	880.4**	32.61	[816.5, 944.3]
Slope									
Pre-[d]									
Low[b]	-25.82**	5.713	[-37.02, -14.62]	5.912	8.132	[-10.03, 21.85]	31.73**	9.674	[12.77, 50.69]
Medium[b]	-21.19**	4.238	[-29.50, -12.88]	-14.26*	6.616	[-27.23, -1.30]	6.926	7.590	[-7.95, 21.80]
High[b]	-21.20**	1.379	[-23.90, -18.49]	0.272	2.473	[-4.58, 5.12]	21.47**	2.736	[16.11, 26.83]

Table 6 (continued).

Parameter	Ontario and British Columbia			Ontario and Newfoundland and Labrador			British Columbia and Newfoundland and Labrador		
	$\hat{B}$	*SE*	95% CI	$\hat{B}$	*SE*	95% CI	$\hat{B}$	*SE*	95% CI
Slope									
Post-[e]									
Low[b]	18.21	12.27	[-5.83, 42.26]	15.73	18.69	[-20.90, 52.36]	-2.485	21.47	[-44.56, 39.59]
Medium[b]	6.167	8.418	[-10.33, 22.67]	9.304	14.98	[-20.06, 38.67]	3.137	16.47	[-29.14, 35.41]
High[b]	13.05**	2.194	[8.75, 17.35]	10.54*	3.787	[3.12, 17.96]	-2.508	4.204	[-10.75, 5.73]

Note. $\hat{B}$ = unstandardized regression coefficient; *SE* = robust standard error; CI = confidence interval.

[a] = weeks 1 – 12; [b] = 0 ≤ 4, 5 ≤ 8, and 9 ≤ 12 weeks the application was opened at least once pre-intervention for low, medium, and high engagement, respectively; [c] = weeks 18 - 25; [d] = week 12; [e] = week 21.

* = p < .05; ** = p <.01.

4.4 *Sensitivity Analyses*

Within and between province estimates of the pre- and post-intervention intercept and slope for the total study sample (users with and without missing data) using multiply imputed data are presented in Tables 7 and 8, respectively. Consistent with the primary analysis, pre- to post-intervention estimated intercept dropped in all three provinces with the most pronounced decrease noted in ON ($\gamma_{12} - \gamma_{21}$ = -159 steps/d, $p < .01$), BC ($\gamma_{12} - \gamma_{21}$ = -89 steps/d, $p < .01$), and NL ($\gamma_{12} - \gamma_{21}$ = -40 steps/d, $p < .01$; Table 7). Furthermore, the pre- to post-intervention intercept difference was greatest when comparing ON to BC and NL (70.2/d and 117.9 steps/d, respectively; Table 8). In terms of weekly mean daily step count slopes, significant differences between pre- and post-intervention estimated slope were observed in ON ($\beta_1 + \beta_{13}$ = 5.941 steps/wk, *SE* = 0.667 [4.63, 7.25], $p < .01$) and BC ($\beta_1 + \beta_{13}$ = -5.551 steps/wk, *SE* = 1.049 [-7.61, -3.50], $p < .01$), but not NL ($\beta_1 + \beta_{13}$ = -2.810 steps/wk, *SE* = 2.989 [-8.67, 3.05]); Table 7). Between provinces analyses show that the *positive* estimated post-intervention slope in ON was significantly different from BC ($\beta_1 + \beta_{13}$ = 11.49 steps/wk, *SE* = 1.243 [9.06, 13.93], $p < .01$) and NL ($\beta_1 + \beta_{13}$ = 8.751 steps/wk, *SE* = 3.063 [2.75, 14.75], $p < .01$; Table 8), though the rate of change was modest in terms of steps/d (1.64 and 1.25, respectively).

Table 7. Estimated weekly mean daily step count intercepts and slopes (within provinces), total sample.

Parameter	Ontario			British Columbia			Newfoundland and Labrador		
	$\hat{B}$	*SE*	95% CI	$\hat{B}$	*SE*	95% CI	$\hat{B}$	*SE*	95% CI
Intercept									
Pre-[a]	6070**	4.062	[6062, 6078]	6319**	7.496	[6304, 6333]	5555**	18.85	[5518, 5592]
Post-[b]	5911**	3.872	[5904, 5919]	6230**	7.920	[6215, 6246]	5515**	17.95	[5479, 5550]
Slope									
Pre-[c]	-48.19**	0.333	[-48.84, -47.54]	-31.79**	0.665	[-33.09, -30.49]	-41.33**	1.598	[-44.46, -38.20]
Post-[d]	5.941**	0.667	[4.63, 7.25]	-5.551**	1.049	[-7.61, -3.50]	-2.810	2.989	[-8.67, 3.05]

Note. $\hat{B}$ = unstandardized regression coefficient; *SE* = robust standard error; CI = confidence interval.

[a] = study week 12; [b] = study week 21; [c] = study weeks 1 - 12; [d] = study weeks 18 - 25.

[*] = $p < .05$; [**] = $p < .01$.

Table 8. Estimated weekly mean daily step count intercepts and slopes (between provinces), total sample.

Parameter	**Ontario and British Columbia**			**Ontario and Newfoundland and Labrador**			**British Columbia and Newfoundland and Labrador**		
	$\hat{B}$	***SE***	**95% CI**	$\hat{B}$	***SE***	**95% CI**	$\hat{B}$	***SE***	**95% CI**
Intercept									
Pre-[a]	-248.7**	8.526	[-265.4, -232.0]	514.7**	19.28	[476.9, 552.4]	763.4**	20.28	[723.6, 803.1]
Post-[b]	-318.9**	8.816	[-336.2, -301.6]	396.8**	18.36	[360.8, 432.8]	715.7**	19.62	[677.2, 754.1]
Slope									
Pre-[c]	-16.40**	0.744	[-17.86, -14.94]	-6.863**	1.632	[-10.06, -3.66]	9.536**	1.731	[6.14, 12.93]
Post-[d]	11.49**	1.243	[9.06, 13.93]	8.751**	3.063	[2.75, 14.75]	-2.741	3.168	[-8.95, 3.47]

Note. $\hat{B}$ = unstandardized regression coefficient; *SE* = robust standard error; CI = confidence interval.

[a] = study week 12; [b] = study week 21; [c] = study weeks 1 - 12; [d] = study weeks 18 - 25.

* = $p < .05$; ** = $p < .01$.

Chapter 5

5 Discussion

5.1 *Main Findings*

Systematic exploration of commercially available PA app data may accelerate scientific advances in the mHealth field and ultimately improve population-level PA (2018 Physical Activity Guidelines Advisory Committee, 2018). This is one of the first population-level QEs to examine the effect of FI removal on PA. Overall, we found that weekly mean daily step count significantly decreased from pre- to post-intervention within all provinces but this decrease was most pronounced in ON when compared to BC and NL (i.e. 198 and 274 fewer steps/d, respectively). In other words, after daily rewards were removed ON users accumulated roughly 15 to 20 fewer minutes of walking *per week* compared to BC and NL. This difference is clinically relevant given lower morbidity and mortality rates observed for adults accumulating, for example, 100 minutes of MVPA per week vs. 80 minutes of MVPA per week (Warburton & Bredin, 2017). Even though reward removal appeared to negatively impact PA in ON, post-intervention weekly mean daily step count was similar to baseline levels (5786 and 5751 steps/d, respectively). We acknowledge, however, that post-intervention PA was assessed during the cold Canadian Winter months as part of this study whereas baseline PA data for ON users was collected throughout the year when PA levels may have been higher (Mitchell, Lau, et al., 2020).

Sensitivity analyses using the total study sample also generally support our main finding though the decrease in weekly mean daily step count from pre- to post-intervention in ON relative to BC and NL was less pronounced (i.e., 70 and 119 steps/d, respectively). In addition, the decline in PA after FI removal appeared to be influenced by level of app engagement and pre-intervention PA. Decreases in PA were greater among highly engaged and physically active users relative to less engaged and sedentary users (i.e., 117 and 339 steps/d, respectively). Conversely, app experience, age, and gender did not appear to influence the daily step count decline in ON as suggested by modest between-level differences (i.e., range of 10 – 53 steps/d). Post-intervention PA rates of change between provinces were similar (i.e., < 1.8 step/d differences). When examined by covariate, differences in post-intervention PA rates of change were slightly greater within and between provinces (i.e., < 28.2 step/d).

5.2 ***Practical Implications***

Future population-level incentive-based PA interventions should consider the potentially negative impact of FI removal on PA, especially among certain subgroups (e.g., more highly engaged users). *Carrot Rewards* was discontinued in June 2019, in large part due to a lack of long-term funding from provincial and territorial governments (Marotta, 2019; Rondina et al., 2020). Interestingly, a recently published cost-effectiveness analysis of *Carrot Rewards* suggests greater cost-effectiveness among more highly engaged users (Rondina et al., 2021). Governments and corporations with ongoing (e.g., National Steps Challenge, Singapore; Yao et al., 2020) or planned (e.g., Health Incentives Scheme, United Kingdom; Department of Health and Social Care & Churchill, 2021) investments in incentive-based mHealth apps for PA should consider avoiding FI removal in response to high user engagement to control costs. Rather, one practical implication from this research might be to encourage governments/corporations to increase user exposure to natural PA reinforcers (i.e., improved mental health; Mammen & Faulkner, 2013) as FIs are gradually removed over time (i.e., schedule thinning; LeBlanc et al., 2002). Such an approach may protect against the often-cited drawback of FI interventions, which is that intrinsic motivation is undermined with FIs and people revert back to baseline behaviours when rewards are removed (Deci et al., 1999; Promberger & Marteau, 2013). Compared to high-frequency reinforcement (i.e., constant FI provision), schedule thinning is more similar to naturally occurring reinforcers that increase the likelihood of maintained treatment effects once an intervention (i.e., FIs) is withdrawn (Stokes & Baer, 1977).

5.3 ***Theoretical Implications***

Our findings have a number of theoretical implications as well. First, our results are generally consistent with the long-standing SDT suggestion that external rewards undermine intrinsic motivations to do enjoyable tasks such as completing puzzles (Deci et al., 1999). More than 50 years of lab-based psychology research suggests that when people are rewarded to engage in interesting tasks they may otherwise enjoy, intrinsic motivation may be “crowded out” by the external driver, damaging the potential for sustained participation (Promberger & Marteau, 2013). High levels of pre-intervention app engagement (9.37 out of 12 weeks), combined with little *a priori* communications regarding FI withdrawal, led to an acute drop in PA in ON. Though self-determined motivation was not directly measured in this study, this observation provides some insight into the degree to which users in general were externally motivated by the incentive-based app.

On the other hand, for sedentary users, a modest but statistically significant increase in PA from the pre- to the post-intervention periods (i.e., 107 steps/d) suggests FIs may not have “crowded out” intrinsic motives in this sub-group. This is consistent with novel findings by Promberger & Marteau (2013) who found no evidence that rewards undermine intrinsic motivation *for health behaviours* for which people often begin with low levels of intrinsic motivation to begin with (Promberger & Marteau, 2013). Cognitive Evaluation Theory, a sub-theory of SDT that defines social/environmental factors that promote intrinsic motivation, suggests that providing external rewards for realistic PA goals may actually foster internalized motives through increases in perceived competence (R. M. Ryan & Deci, 2000). This may be particularly true for sedentary adults who have very low levels of perceived competence and intrinsic motivation for PA (Mcauley et al., 1994). Alternatively, it is also possible that sedentary users were simply less engaged with the app (limiting their FI earnings) minimizing the impact of FI removal on weekly mean daily step count. Finally, regression to the mean could also explain the PA increases and decreases observed for sedentary and physically active users, respectively. Regression to the mean is a statistical phenomenon where measures at extreme ends of a sample distribution regress toward the true mean of the sample population with repeated measurement (Barnett et al., 2005).

5.4 ***Comparison to Existing Literature***

Our findings should be considered in light of similar literature examining PA after FI removal. First, Mitchell, Orstad, et al. (2020) conducted a meta-analysis of the RCT evidence examining short- (< 6 months) and long-term (≥ 6 months) effects of FIs on daily step count. An important secondary objective was to determine whether PA persisted after FI removal. When individual study estimates were pooled, significant differences in daily step count from baseline were observed during the intervention period (i.e., 607 steps/d) and post-intervention follow-up three to six months after FI removal (i.e., 514 steps/d). Participants included in Mitchell, Orstad, et al. (2020) were from RCTs and given earlier notification of FI removal (i.e., ≥ 3 weeks) compared to in this study (i.e., 5 days) which could have contributed to the discrepancy in decreased PA during post-intervention follow-up (i.e., 93 versus 367 steps/d, respectively). However, our results are consistent with the narrative summary from Mitchell, Orstad, et al. (2020), where vote counting indicated that only four of 18 studies with follow-up data reported positive post-intervention effects (Mitchell, Orstad, et al., 2020).

Second, Pope & Harvey (2015) conducted a 24-week RCT to examine the impact of continued and discontinued FIs on intrinsic and extrinsic motives for fitness-center attendance in first-year college students. Participants in the discontinued-incentive conditions met significantly fewer fitness-center attendance goals (3%) relative to the continued-incentive condition (39%). Notably, intrinsic motives were not significantly different over time or by condition. However, in accordance with Attribution Theory (which theorizes that individuals try to explain their behaviour; Heider, 1958), the authors speculated that participants in the FI groups may have attributed their decline in fitness-center attendance after FI removal to the lack of FI provision. If participants associated their decrease in fitness center attendance to lack of FI provision, they may not have attributed the decline to shifts in intrinsic or extrinsic motivation that were reflected in measures of self-determined motivation (Pope & Harvey, 2015).

Next, our findings should also be compared to prior research that has examined the impact of engagement and *ex-ante* level of PA on daily step count after the removal of FIs. Omran et al. (2018) conducted an 11-week RCT to determine whether FI provision increased daily step count through engagement with an action planning tool built-into a web-based walking intervention. Large effect sizes in favour of the FI condition were observed during the post-incentive period for the number of action plans completed and the change in average daily step count from baseline (1793 steps/d; Omran et al., 2018). Omran et al. (2018) measured engagement with a behaviour change component (i.e., number of action plans completed) whereas we measured engagement through user interaction (i.e., number of weeks the app was opened; Cole-Lewis et al., 2019). Therefore, the decline in step count among highly engaged users in our study may be a result of a difference in the operational definition of engagement.

Mason et al. (2018) conducted an 8-week retrospective cohort study to determine the effectiveness of an incentive-based workplace wellness program aimed at increasing daily PA, particularly among the least active employees. Participants were grouped by baseline PA (steps/d) into four groups: < 6000 (I), 6000 to 7999 (II), 8000 to 9999 (III), and ≥ 10,000 (IV). Participants in group 1 had the greatest increase in PA from baseline (1656 steps/d) and the second lowest decrease in PA after FI removal (528 steps/d; Mason et al., 2018). Although step count increased among the least active participants in the present study (i.e., sedentary users: 107 steps/d), pre-intervention PA was assessed during FI provision. The discrepancy between findings from our study and Mason et al. (2018) among the least active participants may be explained by a difference in the operational definition of pre-intervention PA. Baseline step count in our study was calculated when users first downloaded the app (i.e., could have been recorded up to 19 months prior) and could have been influenced by seasonal variation in weather (Merchant et al., 2007). Therefore, the average of weekly mean daily step count from study weeks one to 12 was a more reliable and valid measure of pre-intervention PA than baseline step count.

Last, our findings should be compared to the results of a prospective longitudinal study of the web- and app-based Vitality Active Rewards (Vitality, n.d.) short-term FI program in the United Kingdom (Hajat et al., 2019). Hajat et al. (2019) found that the number of annual weeks which users (n = 11,881) achieved WHO PA recommendations (i.e., ≥ 150 min/wk of MVPA; World Health Organization, 2010) significantly increased by 19% 24-months post-intervention (i.e., 22.2 to 26.4 wks). Furthermore, this increase was greatest among low-active users (i.e., 316%; 4.9 to 15.5 wks) and a small but significant decrease was noted in high-active users (i.e., 2.7%; 40.4 to 39.3 wks). However, achievement of the WHO PA recommendations was calculated on the assumption that each day in which FIs were earned corresponded to at least 30 minutes of MPA or 15 minutes of VPA. Given that FIs were provided for daily PA (i.e., gym visits, step count, and social running events) that may not have equated to WHO PA recommendations, it is not possible to quantify the level in which PA was sustained after the removal of FIs (Hajat et al., 2019). In contrast, PA was objectively measured in our study by weekly mean daily step count which can be conservatively translated to intensity-based guidelines (Tudor-Locke et al., 2013) from the public health organizations (i.e., WHO; World Health Organization, 2020b).

5.5 ***Limitations***

A number of limitations should be considered when interpreting the results of this study. First, since randomization to experimental and control conditions was not possible within this quasi-experimental (i.e., observational) study design the internal validity of our conclusions may be limited. The external validity, however, may be greater than in more carefully controlled RCT studies where internal validity is prioritized. Although baseline (i.e., age) and pre-intervention (i.e., app engagement) covariates were balanced between provinces with regression adjustment to minimize selection bias, we could not minimize the confounding effect of unmeasured variables. For instance, religious differences between provinces could have impacted mean daily step count during the Christmas/holiday season through variations in PA routine and smartphone “wear time”. Christmas is a central celebration to the Christian liturgical year (Forbes, 2007). However, NL exhibited the smallest post-intervention decrease in mean daily step count (i.e., 93 steps/d) despite 93.2 percent of the population identifying as Christian, relative to ON (63.2%) and BC (44.6%; Statistics Canada, 2013). In addition, seasonal variation between provinces could have differentially affected mean daily step count from pre- to post-intervention. The average daily temperature and precipitation during the study period of the largest cities in ON (Toronto), BC (Vancouver), and NL (St. John’s) was 2.1 ℃/2.4 mm, 6.6 ℃/5.1 mm, and -1.9 ℃/4.6 mm, respectively (Environment and Climate Change Canada, n.d.; Statistics Canada, 2017). Although St. John’s recorded the coldest daily temperatures and received a similar amount of daily precipitation to Vancouver (~ two times that of Toronto), it did not appear to augment the decrease in post-intervention PA in NL (i.e., 93 steps/d) relative to BC and ON (i.e., 169 and 367 steps/d, respectively). Finally, the removal of FI in ON could have reduced smartphone "wear time" compared to BC and NL which may have contributed to the greater decline in *measured* PA in ON. However, assessment of smartphone "wear time" is still an active area of research (Duncan et al., 2017) and could not be evaluated. Future research should consider using an established proxy of "wear time" when analyzing step count data in incentive-based PA apps, such as the time between the first and last recorded step each day (Althoff et al., 2017).

Second, if data was missing not at random (MNAR) it would have violated the MAR assumption required for MI and biased the results of the sensitivity analysis. Data that is MNAR is dependent on unobserved variables even after conditioning on observed data (Little & Rubin, 2014). Nevertheless, testing for MAR versus MNAR is not possible (van Buuren, 2018). Furthermore, no standardized method exists nor should be prescribed for conducting a sensitivity analysis to assess the potential impact of departures from the MAR assumption as it is still an ongoing area of research (Carroll et al., 2004).

Third, the number of days included in the calculation of mean weekly step count ranged from one to seven. The average number of days, however, included in mean weekly step count calculations for the total sample was 5.94, 6.04, and 5.86 in ON, BC, and NL, respectively. Additionally, the number of weeks included in the calculation of pre-intervention level of PA ranged from one to 12. However, the average number of weeks with weekly mean daily step count data in the total sample as determined by pre-intervention app engagement (weekly mean daily step count recorded with each app opening) was 8.11, 8.21, and 7.73 in ON, BC, and NL, respectively.

Fourth, we could not assess psychosocial determinants of engagement (i.e., self-determined motivation) that may have moderated the relationship between app engagement and PA after the removal of FI. However, there is no current definition of engagement that is universally acknowledged (Cole-Lewis et al., 2019) which designers of the *Carrot Rewards* app could have used to inform data collection of the psychosocial determinants of engagement.

Fifth, 64.6% of users from the analytic sample in ON were female (63% of the total sample) which limits the generalizability of our findings to the entire Canadian population. However, this is consistent with many other mHealth interventions that have also found the majority of their samples to be female (Harris, 2019; Maher et al., 2014, 2015; J. Ryan et al., 2017).

Sixth, linear regression was performed with a robust sandwich estimator of covariance which specifies a heteroskedastic covariance model, rather than ordinary least squares regression which assumes homoscedastic covariance. A heteroskedastic covariance model does not assume constant variance and uncorrelated measurements between time points which improves the accuracy of SEs of estimated coefficients and CIs for repeated measures data (Fitzmaurice et al., 2011). Relative to robust sandwich estimators of covariance, ordinary least squares regression produces more precise CIs for intercept. However, robust sandwich estimators of covariance generate more accurate SEs for intercept and CIs/SEs for slope than ordinary least squares regression (Westman, 2020). Furthermore, it is more computationally efficient to analyze large repeated measures datasets using simple linear regression with a robust sandwich estimator of covariance than a linear mixed effects model which also assumes a heteroskedastic covariance model (Guillaume et al., 2014).

5.6 *Future directions*

Given the concern that FIs can be prohibitively costly (Rondina et al., 2020), future research should focus on effective strategies of implementation *and* removal in incentive-based PA interventions. To confirm our findings, future RCTs and QEs should compare PA in conditions where FIs have been removed with conditions of continual incentivization. Future studies should ascertain whether specific subgroups (e.g., adults with chronic conditions; Mitchell, Orstad, et al., 2020) and reinforcement schedules (i.e., schedule thinning; LeBlanc et al., 2002) are associated with improved PA after FI removal. Identifying subgroups more likely to experience reduced PA levels after FI removal, along with ways of protecting against this drop (e.g., schedule thinning), could inform tailored reward withdrawal procedures in the future, maximizing program scalability and sustainability. However, methods of FI removal are not included in lists of FI design features (Mitchell et al., 2015) to inform prospective incentive-based PA interventions. Therefore, future work should examine whether the available evidence on potential FI removal strategies (i.e., targeting subgroups and schedule thinning) warrants inclusion in lists of FI design features. Future research should also investigate the acceptability of tailored FI removal among stakeholders responsible for PA intervention implementation (e.g., policymakers) and financing (e.g., general public through taxation; Giles et al., 2015). If stakeholders find an incentive-based intervention to be unacceptable, delivery and uptake will likely be low (Bigsby et al., 2017; Giles et al., 2016). In terms of provision, universal FIs tend to be preferred by stakeholders over targeted FIs for specific populations (Hoskins et al., 2019). However, the acceptability of targeted FI removal for subgroups that may be less likely to experience reduced PA after universal FI provision (i.e., less engaged users) requires further examination.

5.7 *Conclusion*

To address the global physical inactivity pandemic, stakeholders in the public and private sector need to implement sustainable and scalable population-level PA interventions. Incentive-based interventions delivered through smartphone apps can increase PA at the population-level and be cost-effective. However, effective strategies to remove FIs that maintain increases in PA are urgently needed for governments and corporations who are unable to finance incentive-based interventions indefinitely. Our study suggests that removing small FIs from a smartphone PA app can reduce weekly mean daily step count on a population-level. In addition, our results indicate that highly engaged and physically active users may experience a greater decline in PA after the removal of FIs. Given our study's sample size and QE design, these findings may be applicable to governments and corporations with ongoing or planned incentive-based PA interventions delivered through smartphone apps at a population-level.

References

2018 Physical Activity Guidelines Advisory Committee. (2018). *2018 Physical Activity Guidelines Advisory Committee Scientific Report*. https://health.gov/sites/default/files/2019-09/PAG_Advisory_Committee_Report.pdf

Acland, D., & Levy, M. R. (2015). Naiveté, projection bias, and habit formation in gym attendance. *Management Science*, *61*(1), 146–160. https://doi.org/10.1287/mnsc.2014.2091

Adams, J., Giles, E. L., McColl, E., & Sniehotta, F. F. (2014). Carrots, sticks and health behaviours: a framework for documenting the complexity of financial incentive interventions to change health behaviours. *Health Psychology Review*, *8*(3), 286–295. https://doi.org/10.1080/17437199.2013.848410

Adams, M. A., Hurley, J. C., Todd, M., Bhuiyan, N., Jarrett, C. L., Tucker, W. J., Hollingshead, K. E., & Angadi, S. S. (2017). Adaptive goal setting and financial incentives: a 2 × 2 factorial randomized controlled trial to increase adults' physical activity. *BMC Public Health*, *17*(1), 286. https://doi.org/10.1186/s12889-017-4197-8

Althoff, T., Sosič, R., Hicks, J. L., King, A. C., Delp, S. L., & Leskovec, J. (2017). Large-scale physical activity data reveal worldwide activity inequality. *Nature*, *547*(7663), 336–339. https://doi.org/10.1038/nature23018

Amagasa, S., Machida, M., Fukushima, N., Kikuchi, H., Takamiya, T., Odagiri, Y., & Inoue, S. (2018). Is objectively measured light-intensity physical activity associated with health outcomes after adjustment for moderate-to-vigorous physical activity in adults? A systematic review. *International Journal of Behavioral Nutrition and*

Physical Activity, *15*(1). https://doi.org/10.1186/s12966-018-0695-z

Apptentive. (2021). *2021 Mobile App Engagement Benchmark Report*. https://www.apptentive.com/2021-benchmark-report/#download-report

American Psychological Association (n.d.). *Additive Effect.* APA Dictionary of Psychology. https://dictionary.apa.org/additive-effect

Balanda, K. P., & Macgillivray, H. L. (1988). Kurtosis : A critical review. *The American Statistician*, *42*(2), 111–119. https://doi.org/10.2307/2684482

Balshem, H., Helfand, M., Schünemann, H. J., Oxman, A. D., Kunz, R., Brozek, J., Vist, G. E., Falck-Ytter, Y., Meerpohl, J., Norris, S., & Guyatt, G. H. (2011). GRADE guidelines: 3. Rating the quality of evidence. *Journal of Clinical Epidemiology*, *64*(4), 401–406. https://doi.org/10.1016/j.jclinepi.2010.07.015

Bandura, A. (1997). *Self-efficacy: The exercise of control*. Worth. https://doi.org/10.1891/0889-8391.13.2.158

Baranowski, T. (2017). Exergaming: Hope for future physical activity? Or blight on mankind? *Journal of Sport and Health Science*, *6*(1), 44–46. https://doi.org/10.1016/j.jshs.2016.11.006

Barnett, A. G., van der Pols, J. C., & Dobson, A. J. (2005). Regression to the mean: What it is and how to deal with it. *International Journal of Epidemiology*, *34*(1), 215–220. https://doi.org/10.1093/ije/dyh299

Barreira, T. V., Tudor-Locke, C., Champagne, C. M., Broyles, S. T., Johnson, W. D., & Katzmarzyk, P. T. (2013). Comparison of GT3X accelerometer and YAMAX pedometer steps/day in a free-living sample of overweight and obese adults. *Journal of Physical Activity and Health*, *10*(2), 263–270. https://doi.org/10.1123/jpah.10.2.263

Becker, G. S. (1988). 3. A theory of rational addiction. *Accounting for Tastes*, *96*(41), 50–76. https://doi.org/10.4159/9780674020658-004

Becker, G. S. (1992). Habits, addictions, and traditions. *Kyklos*, *45*, 327-345. https://doi.org/10.1111/j.1467-6435.1992.tb02119.x

Bigsby, E., Seitz, H. H., Halpern, S. D., Volpp, K., & Cappella, J. N. (2017). Estimating acceptability of financial health incentives. *Health Education and Behavior*, *44*(4), 513–518. https://doi.org/10.1177/1090198116664072

Bort-Roig, J., Gilson, N. D., Puig-Ribera, A., Contreras, R. S., & Trost, S. G. (2014). Measuring and influencing physical activity with smartphone technology: A systematic review. *Sports Medicine*, *44*(5), 671–686. https://doi.org/10.1007/s40279-014-0142-5

Brazauskas, R., & Logan, B. R. (2016). Observational studies: Matching or regression? *Biology of Blood and Marrow Transplantation*, *22*(3), 557–563. https://doi.org/10.1016/j.bbmt.2015.12.005

Brower, J., LaBarge, M. C., White, L., & Mitchell, M. S. (2020). Examining responsiveness to an incentive-based mobile health app: Longitudinal observational study. *Journal of Medical Internet Research*, *22*(8), e16797. https://doi.org/10.2196/16797

Budworth, L., Prestwich, A., Sykes-Muskett, B., Khatun, K., Ireland, J., Clancy, F., & Conner, M. (2019). A feasibility study to assess the individual and combined effects of financial incentives and monetary contingency contracts on physical activity. *Psychology of Sport and Exercise*, *44*, 42–50. https://doi.org/10.1016/j.psychsport.2019.04.021

Bull, F. C., Al-Ansari, S. S., Biddle, S., Borodulin, K., Buman, M. P., Cardon, G., Carty, C., Chaput, J. P., Chastin, S., Chou, R., Dempsey, P. C., Dipietro, L., Ekelund, U., Firth, J., Friedenreich, C. M., Garcia, L., Gichu, M., Jago, R., Katzmarzyk, P. T., … Willumsen, J. F. (2020). World Health Organization 2020 guidelines on physical activity and sedentary behaviour. *British Journal of Sports Medicine*, *54*(24), 1451–1462. https://doi.org/10.1136/bjsports-2020-102955

Bulmer, M. G. (1967). *Principles of Statistics (2nd ed.)*. Dover Publications, Inc. http://www.ru.ac.bd/stat/wp-content/uploads/sites/25/2019/03/102_06_Bulmer_Principles-of-Statistics.pdf

Burns, R. J., Donovan, A. S., Ackermann, R. T., Finch, E. A., Rothman, A. J., & Jeffery, R. W. (2012). A theoretically grounded systematic review of material incentives for weight loss: Implications for interventions. *Annals of Behavioral Medicine*, *44*(3), 375–388. https://doi.org/10.1007/s12160-012-9403-4

Camerer, C. F., & Loewenstein, G. (2003). *Advances in Behavioral Economics*. Princeton: Princeton University Press. https://doi.org/10.1016/0167-4870(91)90008-h

Carroll, R. J., Davidian, M., Dubin, J., Fitzmaurice, G. M., Kenward, M., Mohlenberghs, G., & Roy, J. (2004). Emerging issues in longitudinal and functional data analysis: Discussion of two important missing data issues. *Statistica Sinica*, *14*(3), 989–1020. https://www.jstor.org/stable/24307408?refreqid=excelsior%3A8e137ec7245c635dcaf9306190ec09b1

Carson, V., Hunter, S., Kuzik, N., Gray, C. E., Poitras, V. J., Chaput, J. P., Saunders, T. J., Katzmarzyk, P. T., Okely, A. D., Connor Gorber, S., Kho, M. E., Sampson, M., Lee, H., & Tremblay, M. S. (2016). Systematic review of sedentary behaviour and health indicators in school-aged children and youth: An update. *Applied Physiology, Nutrition and Metabolism*, *41*(6), S240–S265. https://doi.org/10.1139/apnm-2015-0630

Carson, V., Lee, E. Y., Hewitt, L., Jennings, C., Hunter, S., Kuzik, N., Stearns, J. A., Unrau, S. P., Poitras, V. J., Gray, C., Adamo, K. B., Janssen, I., Okely, A. D., Spence, J. C., Timmons, B. W., Sampson, M., & Tremblay, M. S. (2017). Systematic review of the relationships between physical activity and health indicators in the early years (0-4 years). *BMC Public Health*, *17 (Suppl. 5)*, S854. https://doi.org/10.1186/s12889-017-4860-0

Charness, G., & Gneezy, U. (2008). Incentives to exercise. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.905026

Chastin, S. F. M., Craemer, M. De, Cocker, K. De, Powell, L., Cauwenberg, J. Van, Dall, P., Hamer, M., & Stamatakis, E. (2019). How does light-intensity physical activity associate with adult cardiometabolic health and mortality? Systematic review with meta-analysis of experimental and observational studies. *British Journal of Sports*

Medicine, *53*(6), 370. https://doi.org/10.1136/bjsports-2017-097563

Chokshi, N. P., Adusumalli, S., Small, D. S., Morris, A., Feingold, J., Ha, Y. P., Lynch, M. D., Rareshide, C. A. L., Hilbert, V., & Patel, M. S. (2018). Loss-framed financial incentives and personalized goal-setting to increase physical activity among ischemic heart disease patients using wearable devices: The ACTIVE REWARD randomized trial. *Journal of the American Heart Association*, *7*(12). https://doi.org/10.1161/JAHA.118.009173

Cole-Lewis, H., Ezeanochie, N., & Turgiss, J. (2019). Understanding health behavior technology engagement: pathway to measuring digital behavior change interventions. *JMIR Formative Research*, *3*(4), 1–10. https://doi.org/10.2196/14052

Condliffe, S., Işgın, E., & Fitzgerald, B. (2017). Get thee to the gym! A field experiment on improving exercise habits. *Journal of Behavioral and Experimental Economics* , *70*, 23–32. https://doi.org/10.1016/j.socec.2017.07.007

Couper, M. P., Alexander, G. L., Zhang, N., Little, R. J. A., Maddy, N., Nowak, M. A., McClure, J. B., Calvi, J. J., Rolnick, S. J., Stopponi, M. A., & Johnson, C. C. (2010). Engagement and retention: Measuring breadth and depth of participant use of an online intervention. *Journal of Medical Internet Research*, *12*(4), e52. https://doi.org/10.2196/jmir.1430

Coz, D. R. (1984). Interaction. *International Statistics Review*, *52*(1), 1–24. https://doi.org/10.2307/1403235

Deci, E. L., Ryan, R. M., & Koestner, R. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological*

Bulletin, *125*(6), 627–668. https://doi.org/10.1037/0033-2909.125.6.627

Department of Health and Social Care, & Churchill, J. (2021, July 7). *New pilot to help people exercise more and eat better*. GOV.UK. https://www.gov.uk/government/news/new-pilot-to-help-people-exercise-more-and-eat-better

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: Defining "gamification." *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments, MindTrek 2011*, *March 2014*, 9–15. https://doi.org/10.1145/2181037.2181040

Domin, A., Spruijt-Metz, D., Theisen, D., Ouzzahra, Y., & Vögele, C. (2021). Smartphone-based interventions for physical activity promotion: Scoping review of the evidence over the last 10 years. *JMIR MHealth and UHealth*, *9*(7), e24308. https://doi.org/10.2196/24308

Duncan, M. J., Wunderlich, K., Zhao, Y., & Faulkner, G. (2017). Walk this way: validity evidence of iphone health application step count in laboratory and free-living conditions. *Journal of Sports Sciences*, *36*(15), 1–10. https://doi.org/10.1080/02640414.2017.1409855

Edney, S., Plotnikoff, R., Vandelanotte, C., Olds, T., Bourdeaudhuij, I. De, Ryan, J., & Maher, C. (2017). "Active Team" a social and gamified app-based physical activity intervention: randomised controlled trial study protocol. *BMC Public Health*, *17*(1), 859. https://doi.org/10.1186/s12889-017-4882-7

Edney, S., Ryan, J. C., Olds, T., Monroe, C., Fraysse, F., Vandelanotte, C., Plotnikoff, R., Curtis, R., & Maher, C. (2019). User engagement and attrition in an app-based physical activity intervention: Secondary analysis of a randomized controlled trial. *Journal of Medical Internet Research*, *21*(11), e14645. https://doi.org/10.2196/14645

Ekelund, U., Tarp, J., Steene-Johannessen, J., Hansen, B. H., Jefferis, B., Fagerland, M. W., Whincup, P., Diaz, K. M., Hooker, S. P., Chernofsky, A., Larson, M. G., Spartano, N., Vasan, R. S., Dohrn, I. M., Hagströmer, M., Edwardson, C., Yates, T., Shiroma, E., Anderssen, S. A., & Lee, I. M. (2019). Dose-response associations between accelerometry measured physical activity and sedentary time and all cause mortality: Systematic review and harmonised meta-analysis. *The BMJ*, *366*, 1–10. https://doi.org/10.1136/bmj.l4570

Elliott, A. C., & Woodward, W. A. (2007). *Statistical analysis quick reference guidebook*. SAGE Publications, Inc. https://www.doi.org/10.4135/9781412985949

Environment and Climate Change Canada. (n.d.). *Historical Data*. https://climate.weather.gc.ca/historical_data/search_historic_data_e.html

Eysenbach, G. (2005). The law of attrition. *Journal of Medical Internet Research*, *7*(1), 1–9. https://doi.org/10.2196/jmir.7.1.e11

FatSecret. (n.d.). *Your key to success*. https://www.fatsecret.com/

Feter, N., Santos, T. S. dos, Caputo, E. L., & Silva, M. C. da. (2019). What is the role of smartphones on physical activity promotion? A systematic review and meta-analysis. *International Journal of Public Health*, *64*(5), 679–690. https://doi.org/10.1007/s00038-019-01210-7

Filler, A., Kowatsch, T., Haug, S., Wahle, F., Staake, T., & Fleisch, E. (2015). MobileCoach: A novel open source platform for the design of evidence-based, scalable and low-cost behavioral health interventions: Overview and preliminary evaluation in the public health context. *2015 Wireless Telecommunications Symposium (WTS)*, 1-6. https://doi.org/10.1109/WTS.2015.7117255

Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2011). *Applied longitudinal analysis*. John Wiley & Sons, Inc. https://doi.org/10.1002/9781119513469

Forbes, B. D. (2007). *Christmas: A candid history*. University of California Press. https://books.google.ca/books/about/Christmas.html?id=ap8unt4cP54C&redir_esc=y

Frey, B. S. (2017). Policy consequences of pay-for-performance and crowding-out. *Journal of Behavioral Economics for Policy*, *1*(1), 55–59. https://sabeconomics.org/wordpress/wp-content/uploads/JBEP-1-1-8-F.pdf

Frey, B. S., & Benz, M. (2005). From imperialism to inspiration: A survey of economics and psychology. *SSRN Electronic Journal*, *118*. https://doi.org/10.2139/ssrn.316681

Frey, B. S., & Jegen, R. (2001). Motivation crowding theory. *Journal of Economic Surveys*, *15*(5), 589–611. https://doi.org/10.1111/1467-6419.00150

Füzéki, E., Engeroff, T., & Banzer, W. (2017). Health benefits of light-intensity physical activity: A systematic review of accelerometer data of the National Health and Nutrition Examination Survey (NHANES). *Sports Medicine*, *47*(9), 1769–1793. https://doi.org/10.1007/s40279-017-0724-0

García-Pérez, M. A., & Núñez-Antón, V. V. (2003). Cellwise residual analysis in two-way contingency tables. *Educational and Psychological Measurement*, *63*(5), 825–839. https://doi.org/10.1177/0013164403251280

Giles, E. L., Becker, F., Ternent, L., Sniehotta, F. F., McColl, E., & Adams, J. (2016). Acceptability of financial incentives for health behaviours: A discrete choice experiment. *PLoS ONE*, *11*(6), 1–19. https://doi.org/10.1371/journal.pone.0157403

Giles, E. L., Robalino, S., Sniehotta, F. F., Adams, J., & McColl, E. (2015). Acceptability of financial incentives for encouraging uptake of healthy behaviours: A critical review using systematic methods. *Preventive Medicine*, *73*, 145–158. https://doi.org/10.1016/j.ypmed.2014.12.029

Glasgow, R. E., Vogt, T. M., & Boles, S. M. (1999). Evaluating the public health impact of health promotion interventions: The RE-AIM framework. *American Journal of Public Health*, *89*(9), 1322–1327. https://doi.org/10.2105/AJPH.89.9.1322

Harari, G. M., Lane, N. D., Wang, R., Crosier, B. S., Campbell, A. T., & Gosling, S. D. (2016). Using smartphones to collect behavioral data in psychological science. *Perspectives on Psychological Science*, *11*(6), 838–854. https://doi.org/10.1177/1745691616650285

Government of Australia. (n.d.). *Getting Started.* 10,000 Steps. https://www.10000steps.org.au/

Government of Canada. (n.d.-a). *Coronavirus disease (COVID-19): Prevention and risks*. https://www.canada.ca/en/public-health/services/diseases/2019-novel-coronavirus-infection/prevention-risks.html

Government of Canada. (n.d.-b). *Coronavirus disease (COVID-19)*. https://www.canada.ca/en/public-health/services/diseases/coronavirus-disease-covid-19.html

Guertler, D., Vandelanotte, C., Kirwan, M., & Duncan, M. J. (2015). Engagement and nonusage attrition with a free physical activity promotion program: The case of 10,000 Steps Australia. *Journal of Medical Internet Research*, *17*(7), e176. https://doi.org/10.2196/jmir.4339

Guillaume, B., Hua, X., Thompson, P. M., Waldorp, L., & Nichols, T. E. (2014). Fast and accurate modelling of longitudinal and repeated measures neuroimaging data. *NeuroImage*, *94*, 287–302. https://doi.org/10.1016/j.neuroimage.2014.03.029

Guthold, R., Stevens, G. A., Riley, L. M., & Bull, F. C. (2018). Worldwide trends in insufficient physical activity from 2001 to 2016: a pooled analysis of 358 population-based surveys with 1·9 million participants. *The Lancet Global Health*, *6*(10), e1077–e1086. https://doi.org/10.1016/S2214-109X(18)30357-7

Guthold, R., Stevens, G. A., Riley, L. M., & Bull, F. C. (2020). Global trends in insufficient physical activity among adolescents: a pooled analysis of 298 population-based surveys with 1·6 million participants. *The Lancet Child and Adolescent Health*, *4*(1), 23–35. https://doi.org/10.1016/S2352-4642(19)30323-2

Guyatt, G. H., Oxman, A. D., Vist, G. E., Kunz, R., Falck-Ytter, Y., Alonso-Coello, P., Schünemann, H. J., & GRADE Working Group (2008). GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ (Clinical research ed.)*, *336*(7650), 924–926.

https://doi.org/10.1136/bmj.39489.470347.AD

Hajat, C., Hasan, A., Subel, S., & Noach, A. (2019). The impact of short-term incentives on physical activity in a UK behavioural incentives programme. *Npj Digital Medicine*, *2*(1), 1–6. https://doi.org/10.1038/s41746-019-0164-3

Hales, S., Turner-McGrievy, G., Fahim, A., Freix, A., Wilcox, S., Davis, R. E., Huhns, M., & Valafar, H. (2016). A mixed-methods approach to the development, refinement, and pilot testing of social networks for improving healthy behaviors. *JMIR Human Factors*, *3*(1), 1–12. https://doi.org/10.2196/humanfactors.4512

Hales, S., Turner-McGrievy, G. M., Wilcox, S., Fahim, A., Davis, R. E., Huhns, M., & Valafar, H. (2016). Social networks for improving healthy weight loss behaviors for overweight and obese adults: A randomized clinical trial of the social pounds off digitally (Social POD) mobile app. *International Journal of Medical Informatics*, *94*, 81–90. https://doi.org/10.1016/j.ijmedinf.2016.07.003

Handley, M. A., Lyles, C. R., McCulloch, C., & Cattamanchi, A. (2018). Selecting and improving quasi-experimental designs in effectiveness and implementation research. *Annual Review of Public Health*, *39*(1), 1–21. https://doi.org/10.1146/annurev-publhealth-040617-014128

Harkins, K. A., Kullgren, J. T., Bellamy, S. L., Karlawish, J., & Glanz, K. (2017). A trial of financial and social incentives to increase older adults' walking. *American Journal of Preventive Medicine*, *52*(5), e123–e130. https://doi.org/10.1016/j.amepre.2016.11.011

Harris, M. A. (2019). Maintenance of behaviour change following a community-wide gamification based physical activity intervention. *Preventive Medicine Reports*, *13*, 37–40. https://doi.org/10.1016/j.pmedr.2018.11.009

Heider, F. (1958). *The psychology of interpersonal relations*. John Wiley & Sons, Inc. https://doi.org/10.4324/9780203781159

Hekler, E. B., Buman, M. P., Grieco, L., Rosenberger, M., Winter, S. J., Haskell, W., & King, A. C. (2015). Validation of physical activity tracking via Android smartphones compared to ActiGraph accelerometer: Laboratory-based and free-living validation studies. *JMIR MHealth and UHealth*, *3*(2), e36. https://doi.org/10.2196/mhealth.3505

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, *21*(11), 1539–1558. https://doi.org/10.1002/sim.1186

Hoskins, K., Ulrich, C. M., Shinnick, J., & Buttenheim, A. M. (2019). Acceptability of financial incentives for health-related behavior change: An updated systematic review. *Preventive Medicine*, *126*, 105762. https://doi.org/10.1016/j.ypmed.2019.105762

Hunter, R. F., Brennan, S. F., Tang, J., Smith, O. J., Murray, J., Tully, M. A., Patterson, C., Longo, A., Hutchinson, G., Prior, L., French, D. P., Adams, J., McIntosh, E., & Kee, F. (2016). Effectiveness and cost-effectiveness of a physical activity loyalty scheme for behaviour change maintenance: A cluster randomised controlled trial. *BMC Public Health*, *16*(1), 1–12. https://doi.org/10.1186/s12889-016-3244-1

Johnson, B. T., MacDonald, H. V., Bruneau, M. L., Goldsby, T. U., Brown, J. C., Huedo-

Medina, T. B., & Pescatello, L. S. (2014). Methodological quality of meta-analyses on the blood pressure response to exercise: A review. *Journal of Hypertension*, *32*(4), 706–723. https://doi.org/10.1097/HJH.0000000000000097

Johnston, M. (2016). What more can we learn from early learning theory? the contemporary relevance for behaviour change interventions. *British Journal of Health Psychology*, *21*(1), 1–10. https://doi.org/10.1111/bjhp.12165

Kasim, R. M., & Raudenbush, S. W. (1998). Application of Gibbs sampling to nested variance components models with heterogeneous within-group variance. *Journal of Educational and Behavioral Statistics*, *23*(2), 93–116. https://doi.org/10.3102/10769986023002093

Kelders, S. M., Kok, R. N., Ossebaard, H. C., & Gemert-Pijnen, J. E. W. C. Van. (2012). Persuasive system design does matter: A systematic review of adherence to web-based interventions. *Journal of Medical Internet Research*, *14*(6), e152. https://doi.org/10.2196/jmir.2104

Kim, Y., & Steiner, P. (2016). Quasi-experimental designs for causal inference. *Educational Psychologist*, *51*(3–4), 395–405. https://doi.org/10.1080/00461520.2016.1207177

Kohl, H. W., Craig, C. L., Lambert, E. V., Inoue, S., Alkandari, J. R., Leetongin, G., Kahlmeier, S., Andersen, L. B., Bauman, A. E., Blair, S. N., Brownson, R. C., Bull, F. C., Ekelund, U., Goenka, S., Guthold, R., Hallal, P. C., Haskell, W. L., Heath, G. W., Katzmarzyk, P. T., … Wells, J. C. (2012). The pandemic of physical inactivity: Global action for public health. *The Lancet*, *380*(9838), 294–305.

https://doi.org/10.1016/S0140-6736(12)60898-8

Kohl, L. F. M., Crutzen, R., & De Vries, N. K. (2013). Online prevention aimed at lifestyle behaviors: A systematic review of reviews. *Journal of Medical Internet Research*, *15*(7), 1–13. https://doi.org/10.2196/jmir.2665

Kramer, J.-N., Künzler, F., Mishra, V., Smith, S. N., Kotz, D., Scholz, U., Fleisch, E., & Kowatsch, T. (2020). Which components of a smartphone walking app help users to reach personalized step goals? Results from an optimization trial. *Annals of Behavioral Medicine*, *54*(7), 518–528. https://doi.org/10.1093/abm/kaaa002

Kraus, W. E., Janz, K. F., Powell, K. E., Campbell, W. W., Jakicic, J. M., Troiano, R. P., Sprow, K., Torres, A., & Piercy, K. L. (2019). Daily step counts for measuring physical activity exposure and its relation to health. *Medicine and Science in Sports and Exercise*, *51*(6), 1206–1212. https://doi.org/10.1249/MSS.0000000000001932

Krueger, H., Turner, D., Krueger, J., & Ready, A. E. (2014). The economic benefits of risk factor reduction in Canada: Tobacco smoking, excess weight and physical inactivity. *Canadian Journal of Public Health*, *105*(1), 69–78. https://doi.org/10.17269/cjph.105.4084

Kruskal, W. H., & Wallis, W. A. (1952). Use of Ranks in One-Criterion Variance Analysis. *Journal of the American Statistical Association*, *47*(260), 583–621. https://doi.org/10.1080/01621459.1952.10483441

Ku, P., Hamer, M., Liao, Y., Hsueh, M., & Chen, L. (2020). Device-measured light-intensity physical activity and mortality: A meta-analysis. *Scandinavian Journal of Medicine & Science in Sports*, *30*(1), 13–24. https://doi.org/10.1111/sms.13557

Kullgren, J. T., Harkins, K. A., Bellamy, S. L., Gonzales, A., Tao, Y., Zhu, J., Volpp, K. G., Asch, D. A., Heisler, M., & Karlawish, J. (2014). A mixed-methods randomized controlled trial of financial incentives and peer networks to promote walking among older adults. *Health Education & Behavior*, *41*(1_suppl), 43S-50S. https://doi.org/10.1177/1090198114540464

Kwan, B. M., McGinnes, H. L., Ory, M. G., Estabrooks, P. A., Waxmonsky, J. A., & Glasgow, R. E. (2019). RE-AIM in the real world: Use of the RE-AIM framework for program planning and evaluation in clinical and community settings. *Frontiers in Public Health*, *7*, 1–10. https://doi.org/10.3389/fpubh.2019.00345

Laranjo, L., Ding, Di., Heleno, B., Kocaballi, B., Quiroz, J. C., Tong, H. L., Chahwan, B., Neves, A. L., Gabarron, E., Dao, K. P., Rodrigues, D., Neves, G. C., Antunes, M. L., Coiera, E., & Bates, D. W. (2021). Do smartphone applications and activity trackers increase physical activity in adults? Systematic review, meta-analysis and metaregression. *British Journal of Sports Medicine*, *55*(8), 422–432. https://doi.org/10.1136/bjsports-2020-102892

LeBlanc, L. A., Hagopian, L. P., Maglieri, K. A., & Poling, A. (2002). Decreasing the intensity of reinforcement-based interventions for reducing behavior: Conceptual issues and a proposed model for clinical practice. *The Behavior Analyst Today*, *3*(3), 289–300. https://doi.org/10.1037/h0099991

Little, R. J. A., & Rubin, D. B. (2014). *Statistical Analysis with Missing Data (3rd ed.)*. John Wiley & Sons, Inc. https://doi.org/10.1002/9781119013563

Loewenstein, G., Asch, D. A., & Volpp, K. G. (2013). Behavioral economics holds potential to deliver better results for patients, insurers, and employers. *Health Affairs*, *32*(7), 1244–1250. https://doi.org/10.1377/hlthaff.2012.1163

Luong, M.-L. N., Hall, M., Bennell, K. L., Kasza, J., Harris, A., & Hinman, R. S. (2020). The impact of financial incentives on physical activity: A systematic review and meta-analysis. *American Journal of Health Promotion*, *35*(2), 236 - 249. https://doi.org/10.1177/0890117120940133

Maher, C., Ferguson, M., Vandelanotte, C., Plotnikoff, R., Bourdeaudhuij, I. De, Thomas, S., Nelson-Field, K., & Olds, T. (2015). A web-based, social networking physical activity intervention for insufficiently active adults delivered via Facebook app: Randomized controlled trial. *Journal of Medical Internet Research*, *17*(7), e174. https://doi.org/10.2196/jmir.4086

Maher, C., Lewis, L. K., Ferrar, K., Marshall, S., Bourdeaudhuij, I. De, & Vandelanotte, C. (2014). Are health behavior change interventions that use online social networks effective? A systematic review. *Journal of Medical Internet Research*, *16*(2), e40. https://doi.org/10.2196/jmir.2952

Mammen, G., & Faulkner, G. (2013). Physical activity and the prevention of depression: A systematic review of prospective studies. *American Journal of Preventive Medicine*, *45*(5), 649–657. https://doi.org/10.1016/j.amepre.2013.08.001

Markland, D., & Ingledew, D. K. (1997). The measurement of exercise motives: Factorial validity and invariance across gender of a revised Exercise Motivations Inventory. *British Journal of Health Psychology*, *2*(4), 361–376. https://doi.org/10.1111/j.2044-8287.1997.tb00549.x

Markland, D., & Tobin, V. (2004). A modification to the Behavioural Regulation in Exercise Questionnaire to include an assessment of amotivation. *Journal of Sport and Exercise Psychology*, *26*(2), 191–196. https://doi.org/10.1123/jsep.26.2.191

Marotta, S. (2019, June 19). *Ottawa-backed Carrot Rewards app shutting down after failing to find a buyer*. The Globe and Mail. https://www.theglobeandmail.com/business/article-ottawa-backed-carrot-rewards-app-shutting-down-after-failing-to-find-a/

Marquet, O., Alberico, C., & Hipp, A. J. (2018). Pokémon GO and physical activity among college students. A study using ecological momentary assessment. *Computers in Human Behavior*, *81*, 215–222. https://doi.org/10.1016/j.chb.2017.12.028

Mason, M. R., Ickes, M. J., Campbell, M. S., & Bollinger, L. M. (2018). An incentivized, workplace physical activity intervention preferentially increases daily steps in inactive employees. *American Journal of Health Promotion*, *32*(3), 638–645. https://doi.org/10.1177/0890117117723803

Mcauley, E., Lox, C., Rudolph, D., & Travis, A. (1994). Self-efficacy and intrinsic motivation in exercising middle-aged adults. *The Journal of Applied Gerontology*, *13*(4), 355–370. https://doi.org/10.1177/073346489401300402

McGavock, J., Brunton, N., Klaprat, N., Swanson, A., Pancoe, D., Manley, E., Weerasinghe, A., Booth, G. L., Russell, K., Rosella, L., & Hobin, E. (2019). Walking on water - A natural experiment of a population health intervention to promote physical activity after the winter holidays. *International Journal of*

Environmental Research and Public Health, *16*(19), 1–13. https://doi.org/10.3390/ijerph16193627

McLaughlin, M., Delaney, T., Hall, A., Byaruhanga, J., Mackie, P., Grady, A., Reilly, K., Campbell, E., Sutherland, R., Wiggers, J., & Wolfenden, L. (2021). Associations between digital health intervention engagement, physical activity, and sedentary behavior: Systematic review and meta-analysis. *Journal of Medical Internet Research*, *23*(2). https://doi.org/10.2196/23180

Merchant, A. T., Dehghan, M., & Akhtar-Danesh, N. (2007). Seasonal variation in leisure-time physical activity among Canadians. *Canadian Journal of Public Health*, *98*(3), 203–208. https://doi.org/10.1007/BF03403713

Meyerowitz-Katz, G., Ravi, S., Arnolda, L., Feng, X., Maberly, G., & Astell-Burt, T. (2020). Rates of attrition and dropout in app-based interventions for chronic disease: Systematic review and meta-analysis. *Journal of Medical Internet Research*, *22*(9), e20283. https://doi.org/10.2196/20283

Michie, S., West, R., Campbell, R., Brown, J., & Gainforth, H. (2014). *ABC of behaviour change theories*. Silverback Publishing.

Milkman, K. L., Gromet, D., Ho, H., Kay, J. S., Lee, T. W., Pandiloski, P., Park, Y., Rai, A., Bazerman, M., Beshears, J., Bonacorsi, L., Camerer, C., Chang, E., Chapman, G., Cialdini, R., Dai, H., Eskreis-Winkler, L., Fishbach, A., Gross, J. J., … Duckworth, A. L. (2021). Megastudies improve the impact of applied behavioural science. *Nature*, *600*, 478 - 483. https://doi.org/10.1038/s41586-021-04128-4

Mitchell, M., Goodman, J. M., Alter, D. A., Oh, P. I., & Faulkner, G. E. J. (2015).

Development of the Health Incentive Program Questionnaire (HIP-Q) in a cardiac rehabilitation population. *Translational Behavioral Medicine*, *5*(4), 443–459. https://doi.org/10.1007/s13142-015-0330-3

Mitchell, M., Lau, E., White, L., & Faulkner, G. (2020). Commercial app use linked with sustained physical activity in two Canadian provinces: A 12-month quasi-experimental study. *International Journal of Behavioral Nutrition and Physical Activity*, *17*(1), 24. https://doi.org/10.1186/s12966-020-00926-7

Mitchell, M., Orstad, S. L., Biswas, A., Oh, P. I., Jay, M., Pakosh, M. T., & Faulkner, G. (2020). Financial incentives for physical activity in adults: Systematic review and meta-analysis. *British Journal of Sports Medicine*, *54*(21), 1259–1268. https://doi.org/10.1136/bjsports-2019-100633

Mitchell, M., White, L., Lau, E., Leahey, T., Adams, M. A., & Faulkner, G. (2018). Evaluating the Carrot Rewards app, a population-level incentive-based intervention promoting step counts across two Canadian provinces: Quasi-experimental study. *JMIR MHealth and UHealth*, *6*(9), 1–13. https://doi.org/10.2196/mhealth.9912

Mitchell, M., White, L., Oh, P., Alter, D., Leahey, T., Kwan, M., & Faulkner, G. (2017). Uptake of an incentive-based mHealth app: Process evaluation of the Carrot Rewards app. *JMIR MHealth and UHealth*, *5*(5), e70. https://doi.org/10.2196/mhealth.7323

Mönninghoff, A., Kramer, J. N., Hess, A. J., Ismailova, K., Teepe, G. W., Car, L. T., Müller-Riemenschneider, F., & Kowatsch, T. (2021). Long-term effectiveness of mHealth physical activity interventions: Systematic review and meta-analysis of randomized controlled trials. *Journal of Medical Internet Research*, *23*(4), e26699. https://doi.org/10.2196/26699

Moschetti, D. (2013). Wellness incentives: Can the use of financial incentives and penalties be effective for motivating participation in wellness programs and sustaining health behavior modification? *Seminar Research Paper Series. Paper*, *30*. https://digitalcommons.uri.edu/cgi/viewcontent.cgi?article=1037&context=lrc_paper_series

Mottola, M. F., Davenport, M. H., Ruchat, S. M., Davies, G. A., Poitras, V., Gray, C., Jaramillo, A., Barrowman, N., Adamo, K. B., Duggan, M., Barakat, R., Chilibeck, P., Fleming, K., Forte, M., Korolnek, J., Nagpal, T., Slater, L., Stirling, D., & Zehr, L. (2018). Canadian guideline for physical activity throughout pregnancy. *Journal of Obstetrics and Gynaecology Canada*, *40*(11), 1528–1537. https://doi.org/10.1016/j.jogc.2018.07.001

Moves. (2018, July 2). *Thanks for Using Moves*. https://www.moves-app.com/

Murray, E., White, I. R., Varagunam, M., Godfrey, C., Khadjesari, Z., & McCambridge, J. (2013). Attrition revisited: Adherence and retention in a web-based alcohol trial. *Journal of Medical Internet Research*, *15*(8). https://doi.org/10.2196/jmir.2336

Murray, J. M., French, D. P., Kee, F., Gough, A., Tang, J., & Hunter, R. F. (2020). Mechanisms of physical activity behavior change in an incentive-based intervention: Mediation analysis. *Health Psychology*, *39*(4), 281–297. https://doi.org/10.1037/hea0000849

Murray, J. M., French, D. P., Patterson, C. C., Kee, F., Gough, A., Tang, J., & Hunter, R.

F. (2019). Predicting outcomes from engagement with specific components of an internet-based physical activity intervention with financial incentives: Process analysis of a cluster randomized controlled trial. *Journal of Medical Internet Research*, *21*(4), e11394. https://doi.org/10.2196/11394

Naci, H., & Soumerai, S. B. (2016). History bias, study design, and the unfulfilled promise of pay-for-performance policies in health care. *Preventing Chronic Disease*, *13*(6), 1–7. https://doi.org/10.5888/pcd13.160133

Natural Sciences and Engineering Research Council of Canada. (2017). Guide for applicants: considering equity, diversity and inclusion in your application. https://www.nserc-crsng.gc.ca/_doc/EDI/Guide_for_Applicants_EN.pdf

Newbold, J. W., Rudnicka, A., & Cox, A. (2021). Staying active while staying home: The use of physical activity technologies during life disruptions. *Frontiers in Digital Health*, *3*. https://doi.org/10.3389/fdgth.2021.753115

Ng, G. (2018, December 3). *Carrot Rewards to kill free daily step goal rewards in Ontario starting Dec. 8*. IPhone in Canada. https://www.iphoneincanada.ca/news/carrot-rewards-axe-daily-step-goal-ontario/

Ninantic Inc. (n.d.). *Pokémon G*. https://www.pokemon.com/us/pokemon-video-games/pokemon-go/

Nunan, D., Bankhead, C., & Aronson, J. (2017). *Selection bias*. Catalogue Of Bias. https://catalogofbias.org/biases/selection-bias/

Office of Disease Prevention and Health Promotion. (2015). *Scientific Report of the 2015*

Dietary Guidelines Advisory Committee. https://health.gov/sites/default/files/2019-09/Scientific-Report-of-the-2015-Dietary-Guidelines-Advisory-Committee.pdf

Omran, J., Trinh, L., Arbour-Nicitopoulos, K. P., Mitchell, M. S., & Faulkner, G. E. (2018). Do incentives promote action planning in a web-based walking intervention? *American Journal of Health Behavior*, *42*(4), 13–22. https://doi.org/10.5993/AJHB.42.4.2

Patel, M. S., Asch, D. A., Rosin, R., Small, D. S., Bellamy, S. L., Eberbach, K., Walters, K. J., Haff, N., Lee, S. M., Wesby, L., Hoffer, K., Shuttleworth, D., Taylor, D. H., Hilbert, V., Zhu, J., Yang, L., Wang, X., & Volpp, K. G. (2016). Individual versus team-based financial incentives to increase physical activity: A randomized, controlled trial. *Journal of General Internal Medicine*, *31*(7), 746–754. https://doi.org/10.1007/s11606-016-3627-0

Patel, M. S., Asch, D. A., Rosin, R., Small, D. S., Bellamy, S. L., Heuer, J., Sproat, S., Hyson, C., Haff, N., Lee, S. M., Wesby, L., Hoffer, K., Shuttleworth, D., Taylor, D. H., Hilbert, V., Zhu, J., Yang, L., Wang, X., & Volpp, K. G. (2016). Framing financial incentives to increase physical activity among overweight and obese adults: A randomized, controlled trial. *Annals of Internal Medicine*, *164*(6), 385. https://doi.org/10.7326/m15-1635

Patel, M. S., Volpp, K. G., Rosin, R., Bellamy, S. L., Small, D. S., Fletcher, M. A., Osman-Koss, R., Brady, J. L., Haff, N., Lee, S. M., Wesby, L., Hoffer, K., Shuttleworth, D., Taylor, D. H., Hilbert, V., Zhu, J., Yang, L., Wang, X., & Asch, D. A. (2016). A randomized trial of social comparison feedback and financial incentives to increase physical activity. *American Journal of Health Promotion*, *30*(6), 416–424. https://doi.org/10.1177/0890117116658195

Patel, M. S., Volpp, K. G., Rosin, R., Bellamy, S. L., Small, D. S., Heuer, J., Sproat, S., Hyson, C., Haff, N., Lee, S. M., Wesby, L., Hoffer, K., Shuttleworth, D., Taylor, D. H., Hilbert, V., Zhu, J., Yang, L., Wang, X., & Asch, D. A. (2018). A randomized, controlled trial of lottery-based financial incentives to increase physical activity among overweight and obese adults. *American Journal of Health Promotion*, *32*(7), 1568–1575. https://doi.org/10.1177/0890117118758932

Pearson, E., Prapavessis, H., Higgins, C., Petrella, R., White, L., & Mitchell, M. (2020). Adding team-based financial incentives to the Carrot Rewards physical activity app increases daily step count on a population scale: a 24-week matched case control study. *International Journal of Behavioral Nutrition and Physical Activity*, *17*(1), 139. https://doi.org/10.1186/s12966-020-01043-1

Perski, O., Blandford, A., West, R., & Michie, S. (2017). Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. *Translational Behavioral Medicine*, *7*(2), 254–267. https://doi.org/10.1007/s13142-016-0453-1

Petry, N. M., Andrade, L. F., Barry, D., & Byrne, S. (2013). A randomized study of reinforcing ambulatory exercise in older adults. *Psychology and Aging*, *28*(4), 1164–1173. https://doi.org/10.1037/a0032563

Pew Research Center. (2021, April 7). *Mobile fact sheet*. https://www.pewresearch.org/internet/fact-sheet/mobile/

Pope, L., & Harvey, J. (2015). The impact of incentives on intrinsic and extrinsic motives for fitness-center attendance in college first-year students. *American Journal of Health Promotion*, *29*(3), 192–199. https://doi.org/10.4278/ajhp.140408-quan-135

Prochaska, J. O., & Velicer, W. F. (1997). The transtheoretical model of health behavior change. *American Journal of Health Promotion*, *12*(1), 38–48. https://doi.org/10.4278/0890-1171-12.1.38

Promberger, M., & Marteau, T. M. (2013). When do financial incentives reduce intrinsic motivation? Comparing behaviors studied in psychological and economic literatures. *Health Psychology*, *32*(9), 950–957. https://doi.org/10.1037/a0032727

Public Health Agency of Canada. (2015, July 29). *National Healthy Living Platform: "Carrot Rewards" targets lifestyle improvements*. https://www.canada.ca/en/news/archive/2015/07/national-healthy-living-platform-carrot-rewards-targets-lifestyle-improvements.html

Qiu, S., Cai, X., Jia, L., Sun, Z., Wu, T., Wendt, J., Steinacker, J. M., & Schumann, U. (2021). Does objectively measured light-intensity physical activity reduce the risk of cardiovascular mortality? A meta-analysis. *European Heart Journal - Quality of Care and Clinical Outcomes*, *7*(5), 496 - 504. https://doi.org/10.1093/ehjqcco/qcaa051

Qiu, S., Cai, X., Wu, T., Sun, Z., Guo, H., Kirsten, J., Wendt, J., Steinacker, J. M., & Schumann, U. (2020). Objectively-measured light-intensity physical activity and risk of cancer mortality: a meta-analysis of prospective cohort studies. *Cancer Epidemiology Biomarkers and Prevention*, *29*(5), 1067–1073.

https://doi.org/10.1158/1055-9965.EPI-19-1446

Rabin, M. (2002). A perspective on psychology and economics. *European Economic Review*, *46*, 657–685. https://doi.org/10.1016/S0014-2921(01)00207-0

Reis, R. S., Salvo, D., Ogilvie, D., Lambert, E. V, Goenka, S., & Brownson, R. C. (2016). Scaling up physical activity interventions across the globe: stepping up to larger and smarter approaches to get people moving. *Lancet*, *388*(10051), 1337–1348. https://doi.org/10.1016/S0140-6736(16)30728-0

Research2Guidance. (2017). *mHealth App Economics 2017/2018: Current Status and Future Trends in Mobile Health*. https://research2guidance.com/product/mhealth-economics-2017-current-status-and-future-trends-in-mobile-health/

Romeo, A., Edney, S., Plotnikoff, R., Curtis, R., Ryan, J., Sanders, I., Crozier, A., & Maher, C. (2019). Can smartphone apps increase physical activity? Systematic review and meta-analysis. *Journal of Medical Internet Research*, *21*(3), e12053. https://doi.org/10.2196/12053

Rondina, R., Hong, M., Sarma, S., & Mitchell, M. (2021). Is it worth it? Cost-effectiveness analysis of a commercial physical activity app. *BMC Public Health*, *21*(1), 1–10. https://doi.org/10.1186/s12889-021-11988-y

Rondina, R., Pearson, E. K., Prapavessis, H., White, L., Richard, S., Mitchell, M. S. (2020). Bright spots, physical activity investments that (almost) worked: Carrot Rewards app, driving engagement with pennies a day. *British Journal of Sports Medicine*, *54*(15), 927–929. https://doi.org/10.1136/bjsports-2019-100969

Ross, R., Chaput, J. P., Giangregorio, L. M., Janssen, I., Saunders, T. J., Kho, M. E., Poitras, V. J., Tomasone, J. R., El-Kotob, R., McLaughlin, E. C., Duggan, M., Carrier, J., Carson, V., Chastin, S. F., Latimer-Cheung, A. E., Chulak-Bozzer, T., Faulkner, G., Flood, S. M., Gazendam, M. K., … Tremblay, M. S. (2020). Canadian 24-Hour Movement Guidelines for Adults aged 18-64 years and Adults aged 65 years or older: an integration of physical activity, sedentary behaviour, and sleep. *Applied Physiology, Nutrition, and Metabolism*, *45*(10.2, Suppl.), S57–S102. https://doi.org/10.1139/apnm-2020-0467

Ryan, J., Edney, S., & Maher, C. (2017). Engagement, compliance and retention with a gamified online social networking physical activity intervention. *Translational Behavioral Medicine*, *7*(4), 702–708. https://doi.org/10.1007/s13142-017-0499-8

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*(1), 68–78. https://doi.org/10.1037/0003-066X.55.1.68

Schneider, P. L., Crouter, S. E., & Bassett, D. R. (2004). Pedometer measures of free-living physical activity: Comparison of 13 models. *Medicine and Science in Sports and Exercise*, *36*(2), 331–335. https://doi.org/10.1249/01.MSS.0000113486.60548.E9

Selya, A. S., Rose, J. S., Dierker, L. C., Hedeker, D., & Mermelstein, R. J. (2012). A practical guide to calculating Cohen's f^2, a measure of local effect size, from PROC MIXED. *Frontiers in Psychology*, *3*, 1–6. https://doi.org/10.3389/fpsyg.2012.00111

Shaffer, J. P. (1995). Multiple hypothesis testing. *Annual Review of Psychology*, *46*(1), 561–584. https://doi.org/10.1146/annurev.ps.46.020195.003021

Shankar, B. (2019). *Canadian wellness app Carrot Rewards shuts down due to lack of funding*. MobileSyrup. https://mobilesyrup.com/2019/06/19/carrot-rewards-shuts-down-lack-of-funding/

Shea, B. J., Grimshaw, J. M., Wells, G. A., Boers, M., Andersson, N., Hamel, C., Porter, A. C., Tugwell, P., Moher, D., & Bouter, L. M. (2007). Development of AMSTAR: A measurement tool to assess the methodological quality of systematic reviews. *BMC Medical Research Methodology*, *7*, 1–7. https://doi.org/10.1186/1471-2288-7-10

Sieverink, F., Kelders, S. M., & Gemert-Pijnen, J. E. W. C. van. (2017). Clarifying the concept of adherence to eHealth technology: Systematic review on when usage becomes adherence. *Journal of Medical Internet Research*, *19*(12), e402. https://doi.org/10.2196/jmir.8578

Spring, B., Pellegrini, C., McFadden, H. G., Pfammatter, A. F., Stump, T. K., Siddique, J., King, A. C., & Hedeker, D. (2018). Multicomponent mHealth intervention for large, sustained change in multiple diet and activity risk behaviors: The make better choices 2 randomized controlled trial. *Journal of Medical Internet Research*, *20*(6), 1–13. https://doi.org/10.2196/10528

Statistics Canada. (2013, May 8). *2011 National Household Survey: Data tables*. https://www12.statcan.gc.ca/nhs-enm/2011/dp-pd/dt-td/Index-eng.cfm

Statistics Canada. (2017, February 8). *Population and Dwelling Count Highlight Tables, 2016 Census*. https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/hltfst/pd-pl/index-eng.cfm

Statistics Canada. (2021a, September 1). *Canadian Health Measures Survey: Activity monitor data, 2018 - 2019*. https://www150.statcan.gc.ca/n1/daily-quotidien/210901/dq210901c-eng.htm

Statistics Canada. (2021b, June 22). *Smartphone use and smartphone habits by gender and age group*. https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=2210011501&pickMembers%5B0%5D=3.2&pickMembers%5B1%5D=4.3

Sterne, J. A. C., White, I. R., Carlin, J. B., Spratt, M., Royston, P., Kenward, M. G., Wood, A. M., & Carpenter, J. R. (2009). Multiple imputation for missing data in epidemiological and clinical research: Potential and pitfalls. *BMJ (Online)*, *339*(7713), 157–160. https://doi.org/10.1136/bmj.b2393

Sterne, J., Savović, J., Page, M. J., Elbers, R. G., Blencowe, N. S., Boutron, I., Cates, C. J., Cheng, H. Y., Corbett, M. S., Eldridge, S. M., Emberson, J. R., Hernán, M. A., Hopewell, S., Hróbjartsson, A., Junqueira, D. R., Jüni, P., Kirkham, J. J., Lasserson, T., Li, T., McAleenan, A., … Higgins, J. (2019). RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ (Clinical research ed.)*, *366*, l4898. https://doi.org/10.1136/bmj.l4898

Stokes, T. F., & Baer, D. M. (1977). An implicit technology of generalization. *Journal of Applied Behavior Analysis*, *10*(2), 349–367. https://doi.org/10.1901/jaba.1977.10-349

Strohacker, K., Galarraga, O., & Williams, D. M. (2014). The impact of incentives on exercise behavior: A systematic review of randomized controlled trials. *Annals of Behavioral Medicine*, *48*(1), 92–99. https://doi.org/10.1007/s12160-013-9577-4

Sydow, L. (2021, January 28). *Pumped Up: Health and Fitness App Downloads Rose 30% in a Landmark Year for Mobile Wellness*. App Annie. https://www.appannie.com/en/insights/market-data/health-fitness-downloads-rose-30-percent/

Taylor, K., & Silver, L. (2019, February 5). *Smartphone ownership is growing rapidly around the world, but not always equally.* Pew Research Center. Https://Www.Pewresearch.Org/Global/2019/02/05/Smartphone-Ownership-Is-Growing-Rapidly-Around-the-World-But-Not-Always-Equally/. https://www.pewresearch.org/global/2019/02/05/smartphone-ownership-is-growing-rapidly-around-the-world-but-not-always-equally/

Teixeira, P. J., Carraça, E. V., Markland, D., Silva, M. N., & Ryan, R. M. (2012). Exercise, physical activity, and self-determination theory: A systematic review. *International Journal of Behavioral Nutrition and Physical Activity*, *9*(78). https://doi.org/10.1186/1479-5868-9-78

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving Decisions About Health, Wealth, and Happiness*. Yale University Press.

The Nobel Prize. (2017, October 9). *Press Release: The Prize in Economic Sciences 2017*. https://www.nobelprize.org/prizes/economic-sciences/2017/summary/

Tudor-Locke, C., Craig, C. L., Thyfault, J. P., & Spence, J. C. (2013). A step-defined sedentary lifestyle index: <5000 steps/day. *Applied Physiology, Nutrition, and*

Metabolism, *38*(2), 100–114. https://doi.org/10.1139/apnm-2012-0235

Tudor-Locke, C., Leonardi, C., Johnson, W. D., Katzmarzyk, P. T., & Church, T. S. (2011). Accelerometer steps/day translation of moderate-to-vigorous activity. *Preventive Medicine*, *53*(1–2), 31–33. https://doi.org/10.1016/j.ypmed.2011.01.014

Tudor-locke, C., Sisson, S. B., Lee, S. M., Craig, C. L., & Plotnikoff, R. C. (2006). Evaluation of quality of commercial pedometers. *Canadian Journal of Public Health*, *97*(Suppl. 1), S10–S15. https://doi.org/10.1007/BF03405359

van Buuren, S. (2018). *Flexible Imputation of Missing Data (2nd ed.)*. Taylor & Francis Group.

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, *45*(3), 1–67. https://doi.org/10.18637/jss.v045.i03

Vincent, S. D., & Sidman, C. L. (2003). Determining measurement error in digital pedometers. *Measurement in Physical Education and Exercise Science*, *7*(1), 19–24. https://doi.org/10.1207/S15327841MPEE0701_2

Vitality. (n.d.). *Active Reward Partners*. https://www.vitality.co.uk/rewards/partners/active-rewards/

Voils, C. I., Chang, Y., Crandell, J., Leeman, J., Sandelowski, M., & Maciejewski, M. L. (2012). Informing the dosing of interventions in randomized trials. *Contemporary Clinical Trials*, *33*(6), 1225–1230. https://doi.org/10.1016/j.cct.2012.07.011

Waddington, H., Aloe, A. M., Becker, B. J., Djimeu, E. W., Hombrados, J. G., Tugwell, P., Wells, G., & Reeves, B. (2017). Quasi-experimental study designs series—paper 6: risk of bias assessment. *Journal of Clinical Epidemiology*, *89*, 43–52. https://doi.org/10.1016/j.jclinepi.2017.02.015

Warburton, D. E. R., & Bredin, S. S. D. (2017). Health benefits of physical activity: A systematic review of current systematic reviews. *Current Opinion in Cardiology*, *32*(5), 541–556. https://doi.org/10.1097/HCO.0000000000000437

Weibel, A., Wiemann, M., & Osterloh, M. (2014). A behavioral economics perspective on the overjustification effect. In Gagné. M (Eds.), *The Oxford Handbook of Work Engagement, Motivation, and Self-Determination Theory* (pp. 72-84). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199794911.013.008

Welk, G. J., Differding, J. A., Thompson, R. W., Blair, S. N., Dziura, J., & Hart, P. (2000). The utility of the Digi-Walker step counter to assess daily physical activity patterns. *Medicine and Science in Sports and Exercise*, *32*(9 Suppl.), S481–S488. https://doi.org/10.1097/00005768-200009001-00007

West, D. S., Krukowski, R. A., Finkelstein, E. A., Stansbury, M. L., Ogden, D. E., Monroe, C. M., Carpenter, C. A., Naud, S., & Harvey, J. R. (2020). Adding financial incentives to online group-based behavioral weight control: An RCT. *American Journal of Preventive Medicine*, *59*(2), 237–246. https://doi.org/10.1016/j.amepre.2020.03.015

Westman, V. (2020). A small sample study of some sandwich estimators to handle heteroscedasticity [Bachelor's thesis, Uppsala University]. Uppsala University publications. https://uu.divaportal.org/smash/record.jsf?aq2=%5B%5B%5D%5D&c=104&af=%5B%5D&searchType=LIST_LATEST&sortOrder2=title_sort_asc&query=&language=sv&pid=diva2%3A1516317&aq=%5B%5B%5D%5D&sf=all&aqe=%5B%5D&sortOrder=author_sort_asc&onlyFullText=false&noOfRows=50

White, I. R., & Carlin, J. B. (2010). Bias and efficiency of multiple imputation compared with complete-case analysis for missing covariate values. *Statistics in Medicine*, *29*(28), 2920–2931. https://doi.org/10.1002/sim.3944

Willis Towers Watson. (2020). *Best Practices in Health Care Employer Survey*. https://www.willistowerswatson.com/-/media/WTW/Insights/2020/03/BPS-Survey-Report-2019.pdf?modified=20200323174703

World Health Organization. (2010). *Global recommendations on physical activity for health*. https://www.who.int/publications/i/item/9789241599979

World Health Organization. (2016). *Monitoring and evaluating digital health interventions. A practical guide to conducting research and assessment*. https://www.who.int/reproductivehealth/publications/mhealth/digital-health-interventions/en/

World Health Organization. (2018). *Global action plan on physical activity 2018 - 2030: more active people for a healthier world*. https://apps.who.int/iris/bitstream/handle/10665/272722/9789241514187-eng.pdf

World Health Organization. (2019). *WHO guidelines on physical activity, sedentary behaviour, and sleep for children under 5 years of age*. https://apps.who.int/iris/handle/10665/311664

World Health Organization. (2020a). *WHO guidelines on physical activity and sedentary behaviour: Web annex: Evidence profiles*. https://apps.who.int/iris/handle/10665/336657

World Health Organization. (2020b). *WHO guidelines on physical activity and sedentary behaviour*. https://www.who.int/publications/i/item/9789240015128

Wurst, R., Maliezefski, A., Ramsenthaler, C., Brame, J., & Fuchs, R. (2020). Effects of incentives on adherence to a web-based intervention promoting physical activity: Naturalistic study. *Journal of Medical Internet Research*, *22*(7), e18338. https://doi.org/10.2196/18338

Yao, J., Tan, C. S., Chen, C., Tan, J., Lim, N., & Müller-Riemenschneider, F. (2020). Bright spots, physical activity investments that work: National Steps Challenge, Singapore: A nationwide mHealth physical activity programme. *British Journal of Sports Medicine*, *54*(17), 1047–1048. https://doi.org/10.1136/bjsports-2019-101662

Yardley, L., Spring, B. J., Riper, H., Morrison, L. G., Crane, D. H., Curtis, K., Merchant, G. C., Naughton, F., & Blandford, A. (2016). Understanding and promoting effective engagement with digital behavior change interventions. *American Journal of Preventive Medicine*, *51*(5), 833–842. https://doi.org/10.1016/j.amepre.2016.06.015

Appendices

Appendix A: Ethical Approval

Date: 26 August 2021

To: Professor Marc Mitchell

Project ID: 114790

Study Title: Examining the effect of financial incentive withdrawal on physical activity: A 24-week natural experiment of Carrot Rewards app users

Application Type: Continuing Ethics Review (CER) Form

Review Type: Delegated

REB Meeting Date: 07/Sept/2021

Date Approval Issued: 26/Aug/2021

REB Approval Expiry Date: 02/Sep/2022

Dear Professor Marc Mitchell,

The Western University Research Ethics Board has reviewed the application. This study, including all currently approved documents, has been re-approved until the expiry date noted above.

REB members involved in the research project do not participate in the review, discussion or decision.

Western University REB operates in compliance with, and is constituted in accordance with, the requirements of the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans (TCPS 2); the International Conference on Harmonisation Good Clinical Practice Consolidated Guideline (ICH GCP); Part C, Division 5 of the Food and Drug Regulations; Part 4 of the Natural Health Products Regulations; Part 3 of the Medical Devices Regulations and the provisions of the Ontario Personal Health Information Protection Act (PHIPA 2004) and its applicable regulations. The REB is registered with the U.S. Department of Health & Human Services under the IRB registration number IRB 00000940.

Please do not hesitate to

contact us if you have any

questions. Sincerely,

The Office of Human Research Ethics

Note: This correspondence includes an electronic signature (validation and approval via an online system that is compliant with all regulations).

Appendix B: Equations for the Primary and Secondary Analyses

Equation 1. Simple linear regression model, primary analysis

$$\gamma_t = \beta_0 + \beta_1 T + \beta_2 X_2 + \ldots + \beta_7 X_7 + \beta_8 TX_2 + \ldots + \beta_{13} TX_7 + \beta_{14} Gender_i + \ldots + \beta_{17} Age + \beta_{18} Household\ Income_i + \ldots + \beta_{28} Loyalty\ Rewards\ Program_i + \ldots + \beta_{34} Baseline\ Step\ Count + \beta_{35} Pre\text{-}Intervention\ App\ Experience + \beta_{36} Pre\text{-}Intervention\ App\ Engagement + \beta_{37} Pre\text{-}Intervention\ Level\ of\ Physical\ Activity \quad (1)$$

Where T represented the number of weeks since the start of the study and γ_t was weekly mean daily step count at time t. X_i was a seven-level categorical variable indicating the intervention period (i.e., X_1 = pre-intervention; $X_2 – X_6$ = intervention weeks 13, 14, 15, 16, and 17; X_7 = post-intervention). The regression coefficients β_0, β_1, $\beta_1 + \beta_{8-12}$, and $\beta_1 + \beta_{13}$ represented weekly mean daily step count at $T = 0$ and slope during the pre-intervention, intervention, and post-intervention periods, respectively. The additive effects of all covariates are indicated by β_{14-37}, where categorical covariates were represented by i and the number of levels equaled the sum of regression coefficients assigned to each variable (i.e., $\beta_{14} Gender_1$ = female, $\beta_{15} Gender_2$ = male, $\beta_{16} Gender_3$ = other). Pre-intervention app experience, app engagement and level of PA along with age were included as continuous covariates.

Equation 2. Simple linear regression model, secondary analyses

$$\gamma_t = \beta_0 + \beta_1 T + \beta_2 X_2 + \ldots + \beta_7 X_7 + \beta_8 TX_2 + \ldots + \beta_{13} TX_7 + \beta_{14} Z + \beta_{15} ZX_2 + \ldots + \beta_{20} ZX_7 + \beta_{21} TZ_1 + \beta_{22} TZ_1 X_2 + \ldots + \beta_{27} TZ_1 X_7 + \beta_{28} TZ_2 + \beta_{29} TZ_2 X_2 + \ldots + \beta_{34} TZ_2 X_7 + \beta_{35} TZ_3 + \beta_{36} TZ_3 X_2 + \ldots + \beta_{41} TZ_3 X_7 + \beta_{42} Gender_i + \ldots + \beta_{45} Age + \ldots + \beta_{46} Household\ Income_i + \ldots + \beta_{56} Loyalty\ Rewards\ Program_i + \ldots + \beta_{62} Baseline\ Step\ Count + \beta_{63} Pre\text{-}Intervention\ App\ Engagement + \beta_{64} Pre\text{-}Intervention\ Level\ of\ Physical\ Activity \qquad (2)$$

For illustrative purposes, only the three-way interaction between level of pre-intervention app experience, study week, and intervention period is shown. Where T represented the number of weeks since the start of the study and γ_t was weekly mean daily step count at time t. X_i was a seven-level categorical variable indicating the intervention period (i.e., X_1 = pre-intervention; $X_2 – X_6$ = intervention weeks 13, 14, 15, 16, and 17; X_7 = post-intervention). Z was a categorical covariate where i represented level of app experience (i.e., Z_1 = low, Z_2 = medium, Z_3 = high). β_0 represented weekly mean daily step count at T = 0 whereas slope during the pre-intervention period was defined by β_{21}, β_{28}, and β_{35} for low, medium, and high app experience, respectively. The regression coefficients $\beta_{21} + \beta_{22 – 26}$, $\beta_{28} + \beta_{29 – 33}$, and $\beta_{35} + \beta_{36 – 40}$ represented slope during the intervention periods for low, medium, and high app experience, respectively. Finally, slope during the post-intervention period was defined by $\beta_{21} + \beta_{27}$, $\beta_{28} + \beta_{34}$, and $\beta_{35} + \beta_{41}$ for low, medium, and high app experience, respectively. The additive effects of all covariates are indicated by $\beta_{42 – 64}$, where categorical covariates were represented by i and the number of levels equaled the sum of regression coefficients assigned to each variable (i.e., $\beta_{42} Gender_1$ = female, $\beta_{43} Gender_2$ = male, $\beta_{44} Gender_3$ = other). Pre-intervention app engagement and level of PA along with age and gender were analyzed using separate three-way interaction models with the same structure as Equation 2 for pre-intervention app experience. When pre-intervention app experience, app engagement and level of PA along with age were not examined for a three-way interaction they were included in each model as a continuous covariate.

Appendix C: Baseline Characteristics and Pre-Intervention Behaviours of the Total Study Sample and Users Excluded from the Complete Cases Sample

Table 9. Baseline characteristics, total sample.

Variable	Ontario (n = 438,731)	British Columbia (n = 124,101)	Newfoundland and Labrador (n = 21,928)
Age (mean ± SD)[a]	30.89 ± 15.51*	32.78 ± 15.80$_a$	32.78 ± 15.16$_a$
Gender[b]			
Female	276,240 (63.0%)*	79,611 (64.1%)*	15,579 (71.0%)*
Male	156,233 (35.6%)*	42,185 (34.0%)*	6110 (27.9%)*
Other	6258 (1.4%)*	2305 (1.9%)*	239 (1.1%)*
Household Income (CAD/year)[b]			
< 20,000	33,773 (7.7%)*	4831 (3.9%)*	1467 (6.7%)
20,000 > 40,000	45,997 (10.5%)*	7449 (6.0%)*	1984 (9.0%)
40,000 > 60,000	54,928 (12.5%)*	9643 (7.8%)*	2272 (10.4%)*
60,000 > 80,000	46,020 (10.5%)*	7759 (6.3%)*	1935 (8.8%)*
80,000 > 100,000	37,526 (8.6%)*	6091 (4.9%)*	1773 (8.1%)
100,000 > 150,000	40,476 (9.2%)*	6864 (5.5%)*	2188 (10.0%)*
≥ 150,000	28,637 (6.5%)*	3859 (3.1%)*	1403 (6.4%)*
Didn't Complete Survey	4023 (0.9%)*	416 (0.3%)*	154 (0.7%)
Don't Know	14,107 (3.2%)*	1823 (1.5%)*	493 (2.2%)*
Rather Not Say	52,209 (11.9%)*	9207 (7.4%)*	2560 (11.7%)*
NA	81,035 (18.5%)*	66,159 (53.3%)*	5699 (26.0%)
Loyalty Rewards Program[b]			
Aeroplan® Miles	63,553 (14.5%)*	20,053 (16.2%)*	5891 (26.9%)*
Drop Points	19,512 (4.4%)*	3799 (3.1%)*	979 (4.5%)
More Rewards®	726 (0.2%)*	16,857 (13.6%)*	57 (0.2%)*
Petro-Points™	50,187 (11.4%)*	7913 (6.4%)*	150 (0.7%)*
RBC Rewards®	9152 (2.1%)*	3199 (2.5%)*	341 (1.6%)*
SCENE Points®	295,601 (67.4%)*	72,280 (58.2%)*	14,510 (66.1%)
Baseline step count[c] *(steps/day; mean ± SD)*[a]	5780 ± 3818*	5922 ± 3636*	5283 ± 3435*

Note. *SD* = standard deviation. *CAD* = Canadian dollars. Means sharing a common subscript were not significantly different at $p < .05$ according to the Independent-Samples Kruskal-Wallis test.

Table 9 (continued).

---

[a] = Independent-Samples Kruskal-Wallis Test; [b] = Chi-squared test of independence;

[c] = mean daily step count 14-days prior to Study Week 1.

* = $p < .05$.

Table 10. Pre-intervention behaviours, total sample.

Variable	Ontario (n = 438,731)	British Columbia (n = 124,101)	Newfoundland and Labrador (n = 21,928)
App Engagement[a] *(weeks; mean ± SD)*[b]	8.11 ± 4.50*	8.21 ± 4.49*	7.73 ± 4.60*
App Experience[c] *(months; mean ± SD)*[b]	12.23 ± 5.66*	15.77 ± 8.59*	18.70 ± 8.91*
Level of Physical Activity (weekly mean daily step count; mean ± SD)[b]	6408 ± 2978*	6561 ± 3022*	5786 ± 3033*

Note. *SD* = standard deviation. Means sharing a common subscript were not significantly different at $p < .05$ according to the Independent-Samples Kruskal-Wallis test.

[a] = weeks the app was opened at least once during the pre-intervention period (Study Weeks 1 – 12). [b] = Independent-Samples Kruskal-Wallis Test; [c] = months since "Steps" program enabled prior to Week 12.

* = $p < .05$.

Table 11. Baseline characteristics of users excluded from the complete cases sample.

Variable	Ontario (n = 160,585)	British Columbia (n = 76,691)	Newfoundland and Labrador (n = 9459)
Age (years; mean ± SD)[a]	33.70 ± 12.95*†	34.88 ± 12.94$_{a}^{\dagger}$	34.41 ± 12.43$_{a}^{\dagger}$
Gender[b]			
Female	96,496 (60.1%)*†	47,927 (62.5%)*†	6619 (70.0%)*†
Male	61,868 (38.5%)*†	27,401 (35.7%)*†	2712 (28.7%)*
Other	2221 (0.5%)*	1363 (1.8%)*	128 (1.4%)†
Household Income (CAD/year)[c]			
< 20,000	7877 (4.9%)*†	963 (1.3%)*†	384 (4.1%)†
20,000 > 40,000	11,422 (7.1%)*†	1530 (2.0%)*†	510 (5.4%)†
40,000 > 60,000	12,746 (7.9%)*†	1866 (2.4%)*†	547 (5.8%)†
60,000 > 80,000	10,059 (6.3%)*†	1384 (1.8%)*†	441 (4.7%)†
80,000 > 100,000	8255 (5.1%)*†	1052 (1.4%)*†	399 (4.2%)†
100,000 > 150,000	8897 (5.5%)*†	1123 (1.5%)*†	483 (5.1%)*†
≥ 150,000	5607 (3.5%)*†	538 (0.7%)*†	283 (3.0%)†
Didn't Complete Survey	803 (0.5%)*†	79 (0.1%)*†	33 (0.3%)†
Don't Know	2935 (1.8%)*†	353 (0.5%)*†	127 (1.3%)†
Rather Not Say	10,949 (6.8%)*†	1644 (2.1%)*†	553 (5.8%)†
NA	81,035 (50.5%)*†	66,159 (86.3%)*†	5699 (60.2%)†
Loyalty Rewards Program[d]			
Aeroplan® Miles	22,469 (14.0%)*†	11,458 (14.9%)†	2571 (27.2%)*
Drop Points	6907 (4.3%)*†	2184 (2.8%)*†	381 (4.0%)
More Rewards®	325 (0.2%)*†	12,001 (15.6%)*†	35 (0.4%)*
Petro-Points™	19,943 (12.4%)*†	4924 (6.4%)*	63 (0.7%)*
RBC Rewards®	3428 (2.1%)*	2159 (2.8%)*†	129 (1.4%)*
SCENE Points®	107,513 (67.0%)*†	43,965 (57.3%)*†	6280 (66.4%)*
Baseline step count[e] *(steps/day; mean ± SD)*[a]	5831 ± 3986*	5946 ± 3710*	5252 ± 3367*

Note. *SD* = standard deviation. *CAD* = Canadian dollars. Means sharing a common subscript were not significantly different between provinces at $p < .05$ according to the Independent-Samples Kruskal-Wallis test.

[a] = Independent-Samples Kruskal-Wallis Test; [b] = Chi-squared test of independence;

[c] = mean daily step count over 14-days prior to Study Week 1.

Table 11 (continued).

[*] = $p < .05$ between provinces; [†] = $p < .05$ between analytic sample and excluded participants.

Table 12. Pre-intervention period behaviours of users excluded from the complete cases sample.

Variable	Ontario (n = 160,585)	British Columbia (n = 76,691)	Newfoundland and Labrador (n = 9459)
App Engagement[a] *(weeks; mean ± SD)*[b]	$5.92 \pm 4.86^{*\dagger}$	$7.04 \pm 4.78^{*\dagger}$	$5.56 \pm 4.76^{*\dagger}$
App Experience[c] *(months; mean ± SD)*[b]	$11.41 \pm 5.71^{*\dagger}$	$14.68 \pm 8.49^{*\dagger}$	$19.72 \pm 8.77^{*\dagger}$
Level of Physical Activity (weekly mean daily step count; mean ± SD)[b]	$6310 \pm 2982^{*\dagger}$	$6387 \pm 2924^{*\dagger}$	$5685 \pm 2967^{*\dagger}$

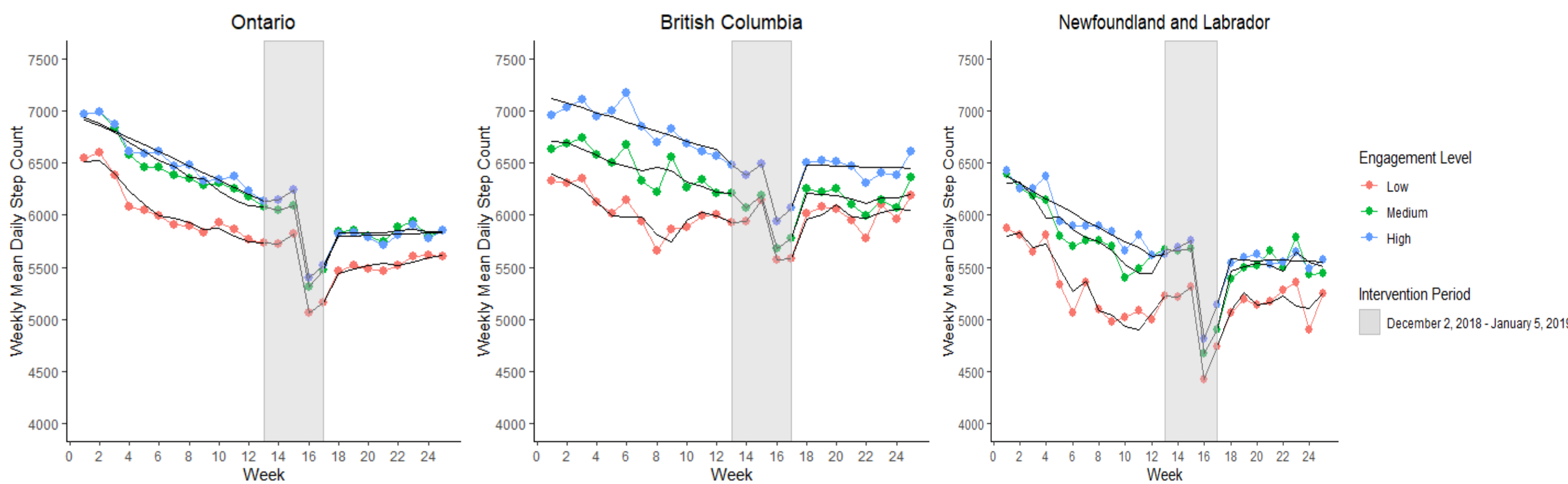
Note. *SD* = standard deviation. Means sharing a common subscript were not significantly different between provinces at $p < .05$ according to the Independent-Samples Kruskal-Wallis test.

[a] = weeks the app was opened at least once during the pre-intervention period (Study Weeks 1 – 12). [b] = Independent-Samples Kruskal-Wallis Test; [c] = months since "Steps" program enabled prior to Week 12.

* = $p < .05$ between provinces; † = $p < .05$ between analytic sample and excluded participants.

Appendix D: Application Engagement, Complete Cases Sample (Secondary Analyses)

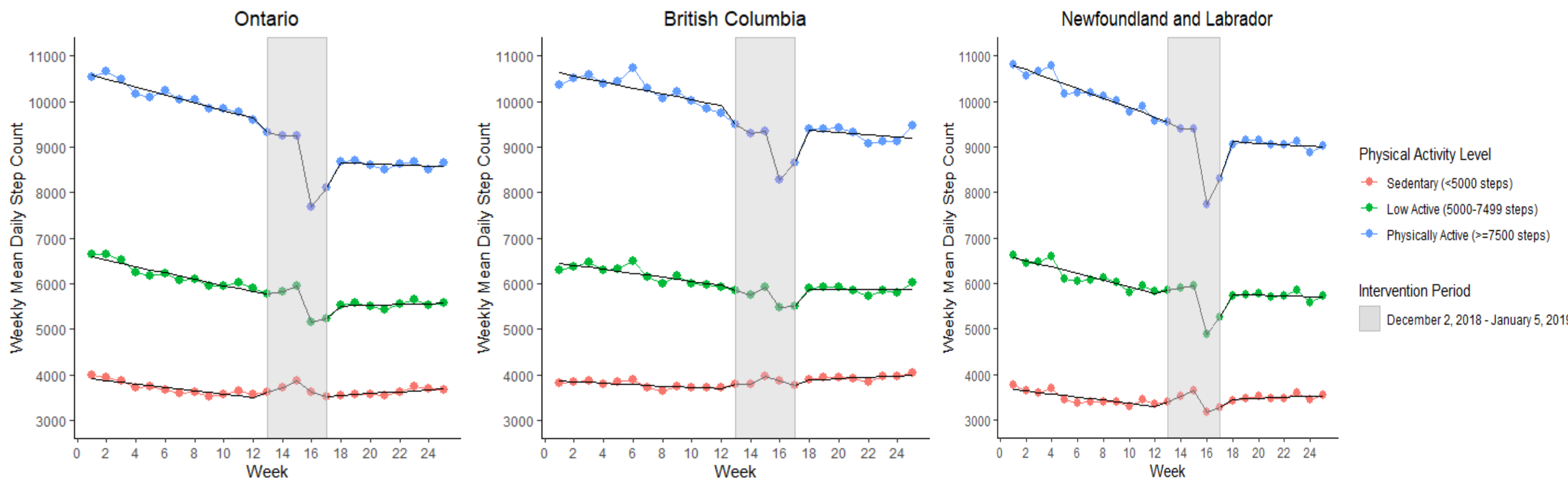
Figure 4. Provincial weekly mean daily step count by week, intervention period, and application engagement.



Note. Observed averages (points) and averages of predictions from the simple linear regression model fit with the three-way interaction between week, intervention period, and level of app engagement (black line) of weekly mean daily step count by week in each province. The intervention period (Study Weeks 13 to 17) from December 2, 2018, to January 5, 2019, was accounted for in the regression model by specifying a separate intervention period level for each of the Weeks 13 to 17. The pre- and post-intervention periods included Weeks 1 -12 and Weeks 18 – 25, respectively. The intervention occurred during Week 13 (December 8, 2018). Engagement level refers to the number of weeks with at least one app opening during the pre-intervention period corresponding to 1 – 4, 5 – 8, and 9 – 12 weeks for low, medium, and high engagement, respectively.

Appendix E: Level of Physical Activity, Complete Cases Sample (Secondary Analyses)

Figure 5. Provincial weekly mean daily step count by week, intervention period, and pre-intervention level of physical activity.



Note. Observed averages (points) and averages of predictions from the simple linear regression model fit with the three-way interaction between week, intervention period, and level of pre-intervention PA (black line) of weekly mean daily step count by week in each province. The intervention period (Study Weeks 13 to 17) from December 2, 2018, to January 5, 2019, was accounted for in the regression model by specifying a separate intervention period level for each of the Weeks 13 to 17. The pre- and post-intervention periods included Weeks 1 -12 and Weeks 18 – 25, respectively. The intervention occurred during Week 13 (December 8, 2018). Physical activity level refers to the average of weekly mean daily step count during the pre-intervention period corresponding to < 5000, 5000 - 7499, and >= (i.e., ≥) 7500 steps for sedentary, low active, and high active levels of pre-intervention physical activity, respectively.

Table 13. Estimated weekly mean daily step count intercepts and slopes by pre-intervention level of physical activity (within provinces), complete cases sample.

Parameter	Ontario			British Columbia			Newfoundland and Labrador		
	$\hat{B}$	*SE*	95% CI	$\hat{B}$	*SE*	95% CI	$\hat{B}$	*SE*	95% CI
Intercept									
Pre-[a]									
Sedentary[b]	3827**	5.658	[3816, 3838]	4025**	13.68	[3998, 4052]	3508**	21.23	[3466, 3549]
Low Active[b]	4504**	4.437	[4496, 4513]	4694**	10.66	[4673, 4715]	4131**	17.81	[4096, 4166]
Physically Active[b]	5469**	4.774	[5460, 5478]	5852**	11.59	[5829, 5874]	4939**	20.41	[4899, 4979]
Post-[c]									
Sedentary[b]	3934**	6.710	[3920, 3947]	4259**	16.30	[4227, 4291]	3695**	25.87	[3644, 3746]
Low Active[b]	4480**	5.183	[4470, 4490]	4818**	12.35	[4794, 4843]	4254**	21.32	[4212, 4295]
Physically Active[b]	5237**	5.700	[5226, 5249]	5805**	13.24	[5779, 5831]	4935**	24.33	[4888, 4983]
Slope									
Pre-[d]									
Sedentary[b]	-38.25**	0.516	[-39.26, -37.24]	-15.89**	1.279	[-18.40, -13.38]	-35.00**	2.096	[-39.11, -30.89]
Low Active[b]	-69.06**	0.777	[-70.58, -67.54]	-43.87**	1.909	[-47.62, -40.13]	-72.91**	4.367	[-58.83, -33.52]
Physically Active[b]	-85.97**	1.189	[-88.31, -83.64]	-65.88**	2.826	[-71.42, -60.34]	-103.3**	6.487	[-116.0, -90.61]

Table 13 (continued).

Parameter	Ontario			British Columbia			Newfoundland and Labrador		
	$\hat{B}$	*SE*	95% CI	$\hat{B}$	*SE*	95% CI	$\hat{B}$	*SE*	95% CI
Slope									
Post-[e]									
Sedentary[b]	24.75**	1.076	[22.64, 26.86]	13.77**	2.669	[8.54, 19.00]	11.23	3.932	[3.53, 18.94]
Low Active[b]	7.934**	1.309	[5.37, 20.50]	-1.65	2.993	[-7.52, 4.22]	-7.398	6.657	[-20.45, 5.65]
Physically Active[b]	-10.98**	1.957	[-14.82, -7.15]	-24.57*	4.226	[-32.85, -16.29]	-18.20	10.09	[-37.97, 1.56]

Note. $\hat{B}$ = unstandardized regression coefficient; *SE* = robust standard error; CI = confidence interval.

[a] = study week 12; [b] = pre-intervention average of weekly mean daily step count of < 5000, 5000 – 7499, and ≥ 7500 for sedentary, low active, and physically active users, respectively; [c] = study week 21; [d] = study weeks 1 – 12; [e] = study weeks 18 - 25.

* = $p < .05$; ** = $p < .01$.

Table 14. Estimated weekly mean daily step count intercepts and slopes by pre-intervention level of physical activity (between provinces), complete cases sample.

Parameter	Ontario and British Columbia			Ontario and Newfoundland and Newfoundland and Labrador			British Columbia and Newfoundland and Labrador		
	$\hat{B}$	*SE*	95% CI	$\hat{B}$	*SE*	95% CI	$\hat{B}$	*SE*	95% CI
Intercept									
Pre-[a]									
Sedentary[b]	-197.6**	14.80	[-226.7, -168.6]	319.6**	21.98	[276.6, 362.7]	517.3**	25.26	[467.8, 566.8]
Low Active[b]	-189.7**	11.54	[-212.4, -167.1]	373.7**	18.35	[337.7, 409.7]	563.4**	20.75	[522.8, 604.1]
Physically Active[b]	-382.6**	12.54	[-407.2, -358.1]	530.5**	20.96	[489.4, 571.6]	913.1**	23.48	[867.1, 959.1]
Post-[c]									
Sedentary[b]	-325.4**	17.62	[-360.0, -290.9]	238.7**	26.73	[186.3, 291.1]	564.1**	30.57	[504.2, 624.0]
Low Active[b]	-338.1**	13.39	[-364.3, -311.8]	226.6**	13.39	[200.4, 252.9]	564.7**	24.64	[516.4, 613.0]
Physically Active[b]	-568.1**	14.42	[-596.3, -539.8]	302.0**	24.99	[253.0, 350.9]	870.0**	27.70	[815.8, 924.3]
Slope									
Pre-[d]									
Sedentary[b]	-22.36**	1.379	[-25.06, -19.66]	-3.246	2.159	[-7.48, -0.99]	19.11**	2.455	[14.30, 23.93]
Low Active[b]	-25.19**	2.061	[-29.23, -21.15]	3.848	4.436	[-4.85, 12.54]	29.04**	4.766	[19.69, 38.38]
Physically Active[b]	-20.10**	3.066	[-26.11, -14.09]	17.35	6.595	[4.42, 30.27]	37.44**	7.076	[23.57, 51.31]

Table 14 (continued).

Parameter	Ontario and British Columbia			Ontario and Newfoundland and Labrador			British Columbia and Newfoundland and Labrador		
	$\hat{B}$	*SE*	95% CI	$\hat{B}$	*SE*	95% CI	$\hat{B}$	*SE*	95% CI
Slope									
Post-[e]									
Sedentary[b]	10.98**	2.878	[5.34, 16.62]	13.52**	4.077	[5.53, 21.51]	2.536	4.752	[-6.78, 11.85]
Low Active[b]	9.584**	3.267	[3.18, 15.99]	15.33*	6.784	[2.04, 28.63]	5.748	7.299	[-8.56, 20.05]
Physically Active[b]	13.59**	4.657	[4.46, 22.71]	7.219	10.27	[-12.92, 27.35]	-6.368	10.94	[-27.80, 15.06]

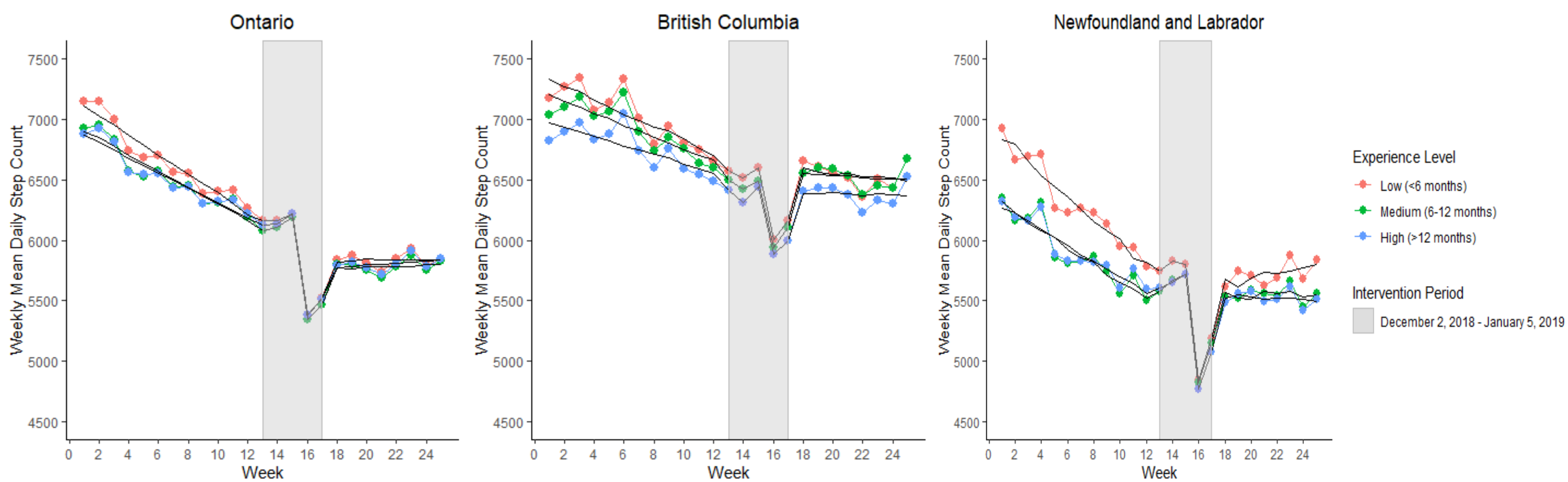
Note. $\hat{B}$ = unstandardized regression coefficient; *SE* = robust standard error; CI = confidence interval.

[a] = study week 12; [b] = pre-intervention average of weekly mean daily step count of < 5000, 5000 – 7499, and ≥ 7500 for sedentary, low active, and physically active users, respectively; [c] = study week 21; [d] = study weeks 1 – 12; [e] = study weeks 18 - 25.

* = $p < .05$; ** = $p < .01$.

Appendix F: Application Experience, Complete Cases Sample (Secondary Analyses)

Figure 6. Provincial weekly mean daily step count by week, intervention period, and application experience.



Note. Observed averages (points) and averages of predictions from the simple linear regression model fit with the three-way interaction between week, intervention period, and level of app experience (black line) of weekly mean daily step count by week in each province. The intervention period (Study Weeks 13 to 17) from December 2, 2018, to January 5, 2019, was accounted for in the regression model by specifying a separate intervention period level for each of the Weeks 13 to 17. The pre- and post-intervention periods included Weeks 1 -12 and Weeks 18 – 25, respectively. The intervention occurred during Week 13 (December 8, 2018). Experience level refers to the number of months prior to week 12 that the “Steps” program was enabled corresponding to < 6, 6 – 12, and > 12 months for low, medium, and high experience, respectively.

Table 15. Estimated weekly mean daily step count intercepts and slopes by application experience (within provinces), complete cases sample.

Parameter	Ontario			British Columbia			Newfoundland and Labrador		
	$\hat{B}$	*SE*	95% CI	$\hat{B}$	*SE*	95% CI	$\hat{B}$	*SE*	95% CI
Intercept									
Pre-[a]									
Low[b]	6064**	7.557	[6050, 6079]	6457**	23.68	[6411, 6504]	5425**	43.80	[5340, 5511]
Medium[b]	6093**	5.466	[6082, 6104]	6473**	18.56	[6436, 6509]	5455**	34.06	[5388, 5521]
High[b]	6125**	4.198	[6117, 6133]	6520**	9.764	[6501, 6539]	5545**	18.66	[5508, 5581]
Post-[c]									
Low[b]	5679**	11.59	[5657, 5702]	6310**	31.44	[6248, 6372]	5285**	64.77	[5158, 5412]
Medium[b]	5712**	8.265	[5696, 5728]	6324**	24.61	[6276, 6372]	5332**	50.14	[5234, 5431]
High[b]	5754**	6.250	[5742, 5767]	6352**	12.87	[6327, 6378]	5432**	26.58	[5380, 5484]
Slope									
Pre-[d]									
Low[b]	-77.75**	1.258	[-80.21, -75.28]	-56.84**	4.139	[-64.96, -48.73]	-90.36**	7.592	[-105.2, -75.48]
Medium[b]	-63.39**	0.816	[-64.99, -61.79]	-46.29**	2.404	[-51.01, -41.58]	-66.38**	4.597	[-75.39, -57.37]
High[b]	-57.87**	0.642	[-59.12, -56.61]	-38.05**	1.458	[-40.91, -35.19]	-56.36**	2.648	[-61.55, -51.17]

Table 15 (continued).

Parameter	Ontario			British Columbia			Newfoundland and Labrador		
	$\hat{B}$	*SE*	95% CI	$\hat{B}$	*SE*	95% CI	$\hat{B}$	*SE*	95% CI
Slope									
Post-[e]									
Low[b]	2.894	2.238	[-1.49, 7.28]	-11.05	6.158	[-23.12, 1.02]	13.51	11.94	[-9.90, 36.91]
Medium[b]	6.416**	1.445	[3.59, 9.25]	-7.260	3.956	[-15.01, 0.49]	-6.314	7.333	[-20.69, 8.06]
High[b]	10.85**	1.119	[8.65, 13.04]	-2.225	2.361	[2.40, -0.94]	-2.847	4.255	[-11.19, 5.49]

Note. $\hat{B}$ = unstandardized regression coefficient; *SE* = robust standard error; CI = confidence interval.

[a] = study week 12; [b] = < 6, 6 ≤ 12, and > 12 months between the date that the "Steps" program was enabled and week 12 for low, medium, and high experience, respectively; [c] = study week 21; [d] = study weeks 1 - 12; [e] = study weeks 18 - 25.

* = $p < .05$; ** = $p < .01$.

Table 16. Estimated weekly mean daily step count intercepts and slopes by application experience (between provinces), complete cases sample.

Parameter	Ontario and British Columbia			Ontario and Newfoundland and Labrador			British Columbia and Newfoundland and Labrador		
	$\widehat{B}$	*SE*	95% CI	$\widehat{B}$	*SE*	95% CI	$\widehat{B}$	*SE*	95% CI
Intercept									
Pre-[a]									
Low[b]	-393.0**	24.86	[-441.7, -344.3]	639.0**	44.45	[551.9, 726.1]	1032**	49.80	[934.4, 1130]
Medium[b]	-379.9**	19.35	[-417.8, -342.0]	638.1**	34.50	[570.5, 705.8]	1018**	38.79	[942.0, 1094]
High[b]	-395.1**	10.63	[-416.0, -374.3]	580.4**	19.12	[542.9, 617.8]	975.5**	21.06	[934.2, 1017]
Post-[c]									
Low[b]	-630.7**	33.51	[-696.4, -565.0]	394.3**	65.80	[265.4, 523.3]	1025**	72.00	[883.9, 1166]
Medium[b]	-612.1**	25.96	[-663.0, -561.2]	379.4**	50.81	[279.8, 479.0]	991.5**	55.85	[882.0, 1101]
High[b]	-598.1**	14.30	[-626.2, -570.1]	322.0**	27.30	[268.5, 375.5]	920.1**	29.53	[862.3, 978.0]
Slope									
Pre-[d]									
Low[b]	-20.90**	4.326	[-29.38, -12.42]	12.62	7.696	[-2.47, 27.70]	33.52**	8.647	[16.57, 50.47]
Medium[b]	-17.10**	2.539	[-22.07, -12.12]	2.991	4.669	[-6.16, -12.14]	20.09**	5.188	[9.92, 30.26]
High[b]	-19.82**	1.593	[-22.94, -16.70]	-1.505	2.725	[-6.85, 3.84]	18.31**	3.023	[12.39, 24.24]

Table 16 (continued).

Parameter	Ontario and British Columbia			Ontario and Newfoundland and Labrador			British Columbia and Newfoundland and Labrador		
	$\hat{B}$	*SE*	95% CI	$\hat{B}$	*SE*	95% CI	$\hat{B}$	*SE*	95% CI
Slope									
Post-[e]									
Low[b]	13.95[*]	6.552	[1.11, 26.79]	-10.61	12.15	[-34.42, 13.20]	-24.56	13.44	[-50.89, 1.78]
Medium[b]	13.68[**]	4.212	[5.42, 21.93]	12.73	7.474	[-1.92, 27.38]	-0.946	8.332	[-17.28, 15.38]
High[b]	13.07[**]	2.613	[7.95, 18.19]	13.69[**]	4.400	[5.07, 22.32]	0.622	4.866	[-8.92, 10.16]

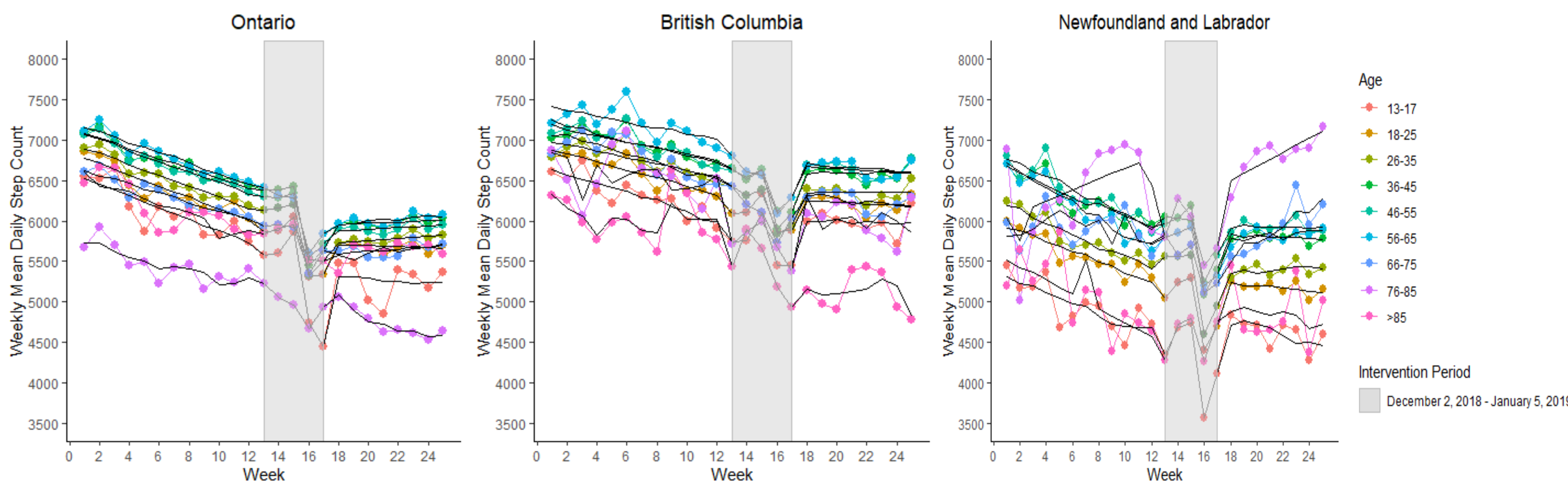
Note. $\hat{B}$ = unstandardized regression coefficient; *SE* = robust standard error; CI = confidence interval.

[a] = study week 12; [b] = < 6, 6 ≤ 12, and > 12 months between the date that the "Steps" program was enabled and week 12 for low, medium, and high experience, respectively; [c] = study week 21; [d] = study weeks 1 - 12; [e] = study weeks 18 - 25.

[*] = $p < .05$; [**] = $p < .01$.

Appendix G: Age, Complete Cases Sample (Secondary Analyses)

Figure 7. Provincial weekly mean daily step count by week, intervention period and age.



Note. Observed averages (points) and averages of predictions from the simple linear regression model fit with the three-way interaction between week, intervention period, and age cohort (black line) of weekly mean daily step count by week in each province. The intervention period (Study Weeks 13 to 17) from December 2, 2018, to January 5, 2019, was accounted for in the regression model by specifying a separate intervention period level for each of the Weeks 13 to 17. The pre- and post-intervention periods included Weeks 1 -12 and Weeks 18 – 25, respectively. The intervention occurred during Week 13 (December 8, 2018).

Table 17. Estimated weekly mean daily step count intercepts and slopes by age (within provinces), complete cases sample.

Parameter	Ontario			British Columbia			Newfoundland and Labrador		
	$\hat{B}$	*SE*	95% CI	$\hat{B}$	*SE*	95% CI	$\hat{B}$	*SE*	95% CI
Intercept									
Pre-[a]									
13 - 17	6105**	11.11	[6084, 6127]	6415**	35.12	[6346, 6483]	5520**	49.84	[5422, 5618]
18 - 25	6112**	8.233	[6096, 6128]	6431**	28.04	[6376, 6486]	5536**	39.88	[5458, 5614]
26 - 35	6118**	7.895	[6103, 6134]	6453**	24.31	[6405, 6500]	5533**	36.29	[5462, 5604]
36 – 45	6120**	9.164	[6102, 6138]	6447**	28.22	[6391, 6502]	5541**	39.16	[5464, 5617]
46 – 55	6113**	9.828	[6094, 6132]	6430**	30.42	[6371, 6490]	5525**	42.72	[5441, 5609]
56 – 65	6110**	10.48	[6089, 6130]	6431**	32.13	[6368, 6494]	5516**	46.74	[5425, 5608]
66 – 75	6106**	10.97	[6085, 6128]	6416**	34.36	[6348, 6483]	5523**	49.22	[5427, 5620]
76 – 85	6105**	11.10	[6084, 6127]	6415**	35.05	[6346, 6484]	5520**	49.81	[5423, 5618]
> 85	6105**	11.10	[6084, 2127]	6415**	35.09	[6346, 6484]	5520**	49.80	[5422, 5618]
Post-[b]									
13 - 17	5617**	16.48	[5585, 5650]	6367**	57.13	[6255, 6479]	5510**	80.68	[5351, 5668]
18 - 25	5677**	12.15	[5653, 5700]	6370**	45.41	[6281, 6459]	5502**	63.73	[5377, 5627]
26 - 35	5671**	11.67	[5648, 5694]	6362**	39.30	[6285, 6439]	5507**	58.43	[5392, 5621]
36 – 45	5652**	13.60	[5625, 5679]	6373**	45.84	[6283, 6463]	5502**	63.05	[5378, 5625]
46 – 55	5633**	14.59	[5604, 5661]	6374**	49.44	[6277, 6470]	5498**	68.94	[5363, 5634]
56 – 65	5622**	15.56	[5591, 5652]	6358**	52.25	[6256, 6460]	5505**	75.63	[5357, 5654]
66 – 75	5619**	16.28	[5587, 5651]	6364**	55.89	[6254, 6473]	5515**	79.68	[5359, 5671]
76 - 85	5618**	16.47	[5585, 5650]	6366**	57.02	[6255, 6478]	5510**	80.64	[5352, 5668]
> 85	5618**	16.48	[5585, 5650]	6366**	57.09	[6254, 6478]	5510**	80.63	[5352, 5668]

Table 17 (continued).

Parameter	Ontario			British Columbia			Newfoundland and Labrador		
	$\widehat{B}$	*SE*	95% CI	$\widehat{B}$	*SE*	95% CI	$\widehat{B}$	*SE*	95% CI
Slope									
Pre-[c]									
13 - 17	-66.21**	1.902	[-69.94, -62.48]	-54.00**	6.050	[-65.86, -42.15]	-62.44**	8.239	[-78.59, -46.29]
18 - 25	-65.07**	0.850	[-66.74, -63.41]	-46.83**	2.464	[-51.66, -42.00]	-53.60**	4.342	[-62.11, -45.09]
26 - 35	-62.23**	0.876	[-64.95, -61.52]	-40.68**	2.158	[-44.91, -36.45]	-62.11**	3.998	[-69.95, -54.28]
36 – 45	-58.85**	1.068	[-60.94, -56.75]	-36.60**	2.491	[-41.49, -31.72]	-60.49**	4.536	[-69.38, -51.60]
46 – 55	-62.30**	1.390	[-65.02, -59.57]	-45.57**	3.402	[-52.24, -38.90]	-69.88**	6.301	[-82.23, -57.53]
56 – 65	-62.78**	2.146	[-66.98, -58.57]	-35.65**	4.600	[-44.67, -26.64]	-83.76**	9.920	[-103.2, -64.31]
66 – 75	-52.95**	4.411	[-61.60, -44.31]	-55.01**	8.176	[-71.03, -38.99]	-26.12	19.77	[-64.87, 12.64]
76 – 85	-40.04**	12.55	[-64.64, -15.44]	-13.27	32.44	[-76.85, 50.32]	68.36	73.77	[-76.21, 212.9]]
> 85	-34.72	18.11	[-70.21, 0.77]	-15.20	32.96	[-79.79, 49.40]	-85.88	54.61	[-192.9, 21.16]
Post-[d]									
13 - 17	-10.02*	3.747	[-17.37, -2.68]	-5.609	11.15	[-27.45, 16.24]	-37.98*	15.81	[-68.98, -6.98]
18 - 25	-3.025	1.568	[-6.10, 0.05]	-3.981	4.262	[-12.34, 4.37]	-20.24*	7.360	[-34.67, -5.82]
26 - 35	18.75**	1.524	[15.77, 21.74]	-0.906	3.404	[-7.58, 5.77]	11.27	6.665	[-1.80, 24.33]
36 – 45	9.671**	1.850	[6.04, 13.30]	2.708	4.057	[-5.24, 10.66]	-2.049	7.136	[-16.04, 11.94]
46 – 55	9.476**	2.326	[4.92, 14.03]	-10.86	5.235	[-21.12, -0.60]	-7.537	9.037	[-25.25, 10.18]
56 – 65	14.37**	3.532	[7.45, 21.30]	-16.27	7.160	[-30.30, -2.23]	10.03	14.07	[-17.55, 37.62]
66 – 75	11.29	7.322	[-3.064, 25.64]	-26.21	13.03	[-51.74, -0.68]	75.27	42.53	[-8.08, 158.6]
76 – 85	-49.93	26.14	[-101.2, 1.31]	-33.59	42.22	[-116.3, 49.17]	87.77	107.6	[-123.2, 298.7]
> 85	6.821	28.50	[-49.04, 62.68]	23.93	45.29	[-64.83, 112.7]	-52.03	44.97	[-140.2, 36.11]

Note. $\widehat{B}$ = unstandardized regression coefficient; *SE* = robust standard error; CI = confidence interval.

[a] = study week 12; [b] = study week 21; [c] = study weeks 1 – 12; [d] = study weeks 18 – 25.

Table 17 (continued).

---

* = p < .05; ** = p < .01.

Table 18. Estimated weekly mean daily step count intercepts and slopes by age (between provinces), complete cases sample.

Parameter	Ontario and British Columbia			Ontario and Newfoundland and Labrador			British Columbia and Newfoundland and Labrador		
	$\hat{B}$	*SE*	95% CI	$\hat{B}$	*SE*	95% CI	$\hat{B}$	*SE*	95% CI
Intercept									
Pre-[a]									
13 - 17	-309.3**	36.83	[-384.4, -237.2]	585.4**	51.06	[485.3, 685.4]	894.6**	60.97	[775.1, 1014]
18 - 25	-318.4**	29.22	[-375.7, -261.2]	575.9**	40.72	[496.1, 655.7]	894.4**	48.74	[798.8, 989.9]
26 - 35	-334.6**	25.26	[-384.7, -284.5]	584.8**	37.14	[512.0, 657.6]	919.4**	43.68	[833.8, 1005]
36 – 45	-326.3**	29.67	[-384.5, -268.2]	579.7**	40.22	[500.8, 658.5]	906.0**	48.27	[811.4, 1001]
46 – 55	-317.5**	31.97	[-380.1, -254.8]	587.9**	43.83	[502.0, 673.8]	905.3**	52.44	[802.6, 1008]
56 – 65	-321.4**	33.80	[-387.6, -255.1]	593.4**	47.90	[499.5, 687.2]	914.7**	56.72	[803.6, 1026]
66 – 75	-309.1**	36.07	[-379.8, -238.4]	583.1**	50.43	[484.2, 681.9]	892.2**	60.03	[774.5, 1010]
76 – 85	-309.7**	36.76	[-381.7, -237.6]	585.0**	51.03	[485.0, 685.0]	894.7**	60.91	[775.3, 1014]
> 85	-309.3**	36.81	[-381.5, -237.2]	585.6**	51.03	[485.6, 685.6]	894.9**	60.93	[775.5, 1014]
Post-[b]									
13 - 17	-749.4**	59.46	[-866.0, -632.9]	107.8	82.35	[-53.63, 269.2]	857.2**	98.86	[663.4, 1051]
18 - 25	-693.1**	47.00	[-785.2, -600.9]	174.8*	64.88	[47.58, 301.9]	867.8**	78.25	[714.4, 1021]
26 - 35	-691.3**	40.99	[-771.6, -610.9]	164.3*	59.59	[47.48, 281.1]	855.6**	70.42	[717.5, 993.6]
36 – 45	-721.0**	47.81	[-814.8, -627.3]	150.3*	64.50	[23.93, 276.8]	871.4**	77.95	[718.6, 1024]
46 – 55	-740.7**	51.55	[-841.7, -639.7]	134.4	70.46	[-3.72, 272.5]	875.1**	84.83	[708.8, 1041]
56 – 65	-736.5**	54.52	[-843.4, -629.7]	116.0	77.21	[-35.30, 267.4]	852.6**	91.92	[672.4, 1033]
66 – 75	-745.0**	58.21	[-859.1, -630.9]	104.1	81.32	[-55.32, 263.5]	849.0**	97.33	[658.3, 1040]
76 - 85	-748.9**	59.35	[-865.2, -632.5]	107.4	82.31	[-53.91, 268.7]	856.3**	98.77	[662.7, 1050]

Table 18 (continued).

Parameter	Ontario and British Columbia			Ontario and Newfoundland and Labrador			British Columbia and Newfoundland and Labrador		
	$\hat{B}$	*SE*	95% CI	$\hat{B}$	*SE*	95% CI	$\hat{B}$	*SE*	95% CI
Intercept									
Post-[b]									
> 85	-748.8**	59.42	[-865.2, -623.3]	108.1	82.29	[-53.23, 269.3]	856.8**	98.79	[663.2, 1050]
Slope									
Pre-[c]									
13 - 17	12.21	6.342	[-24.64, 0.22]	-3.770	8.456	[-20.34, 12.80]	8.438	10.22	[-11.60, 28.47]
18 - 25	-18.24**	2.606	[-23.35, -13.13]	-11.47*	4.424	[-20.14, -2.80]	6.771	4.992	[-3.01, 16.56]
26 - 35	-22.55**	2.329	[-27.12, -17.99]	-1.120	4.093	[-9.14, 6.90]	21.43**	4.543	[12.53, 30.34]
36 – 45	-22.24**	2.710	[-27.55, -16.93]	1.642	4.660	[-7.49, 10.78]	23.88**	5.175	[13.74, 34.03]
46 – 55	-16.73**	3.675	[-23.93, -9.52]	7.585	6.452	[-5.06, 20.23]	24.31**	7.161	[10.28, 38.35]
56 – 65	-27.12**	5.076	[-37.07, 17.18]	20.98*	10.15	[1.09, 40.87]	48.10**	10.94	[26.67, 69.53]
66 – 75	2.059	9.290	[-16.15, 20.27]	-26.84	20.26	[-66.54, 12.87]	-28.89	21.40	[-70.83, 13.04]
76 – 85	-26.77	34.79	[-94.95, 41.41]	-108.4	74.83	[-255.1, 38.25]	-81.63	80.58	[-239.6, 76.31]
> 85	-19.52	37.60	[-93.23, 54.18]	51.16	57.54	[-61.61, 163.9]	70.68	63.79	[-54.34, 195.7]
Post-[d]									
13 - 17	-4.413	11.76	[-27.46, 18.63]	27.96	16.25	[-3.90, 59.81]	32.37	19.35	[-5.55, 70.29]
18 - 25	0.956	4.541	[-7.95, 9.86]	17.22*	7.525	[2.47, 31.97]	16.26	8.505	[-0.41, 32.93]
26 - 35	19.66**	3.730	[12.35, 26.97]	7.483	6.837	[-5.92, 20.88]	-12.17	7.484	[-26.84, 2.49]
36 – 45	6.963	4.459	[-1.78, 15.70]	11.72	7.372	[-2.73, 26.17]	4.757	8.209	[-11.33, 20.85]
46 – 55	20.33**	5.728	[9.11, 31.56]	17.01	9.332	[-1.28, 35.30]	-3.320	10.44	[-23.79, 17.15]
56 – 65	30.64**	7.984	[14.99, 46.28]	4.341	14.51	[-24.10, 32.78]	-26.30	15.79	[-57.24, 4.65]
66 – 75	37.50*	14.94	[8.21, 66.78]	-63.99	43.15	[-148.6, 20.59]	-101.5*	44.48	[-188.7, -14.31]
76 – 85	-16.34	49.66	[-113.7, 80.98]	-137.7	110.8	[-354.8, 79.37]	-121.4	115.6	[-347.9, 105.2]

Table 18 (continued).

Parameter	Ontario and British Columbia			Ontario and Newfoundland and Labrador			British Columbia and Newfoundland and Labrador		
	$\hat{B}$	*SE*	95% CI	$\hat{B}$	*SE*	95% CI	$\hat{B}$	*SE*	95% CI
Post-[d]									
> 85	-17.11	53.51	[-122.0, 87.77]	58.85	53.24	[-45.50, 163.2]	75.97	63.83	[-49.13, 201.1]

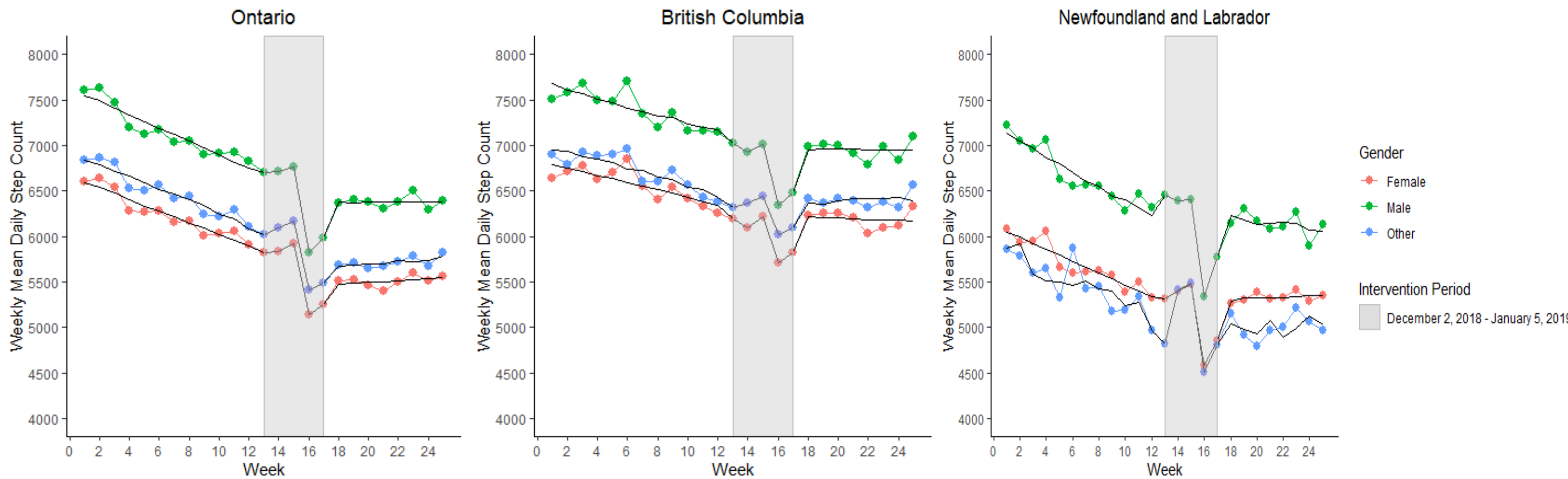
Note. $\hat{B}$ = unstandardized regression coefficient; *SE* = robust standard error; CI = confidence interval.

[a] = study week 12; [b] = study week 21; [c] = study weeks 1 – 12; [d] = study weeks 18 – 25.

* = $p < .05$; ** = $p < .01$.

Appendix H: Gender, Complete Cases Sample (Secondary Analyses)

Figure 8. Provincial weekly mean daily step count by week, intervention period and gender.



Note. Observed averages (points) and averages of predictions from the simple linear regression model fit with the three-way interaction between week, intervention period, and gender subgroup (black line) of weekly mean daily step count by week in each province. The intervention period (Study Weeks 13 to 17) from December 2, 2018, to January 5, 2019, was accounted for in the regression model by specifying a separate intervention period level for each of the Weeks 13 to 17. The pre- and post-intervention periods included Weeks 1 -12 and Weeks 18 – 25, respectively. The intervention occurred during Week 13 (December 8, 2018). Other refers to an identified gender that was not female or male.

Table 19. Estimated weekly mean daily step count intercepts and slopes by gender (within provinces), complete cases sample.

Parameter	**Ontario**			**British Columbia**			**Newfoundland and Labrador**		
	$\hat{B}$	***SE***	**95% CI**	$\hat{B}$	***SE***	**95% CI**	$\hat{B}$	***SE***	**95% CI**
Intercept									
Pre-[a]									
Female	6152**	3.537	[6145, 6159]	6528**	8.574	[6511, 6545]	5583**	15.72	[5552, 5614]
Male	6153**	3.062	[6147. 6159]	6535**	7.331	[6521, 6550]	5574**	14.06	[5546, 5602]
Other[b]	6152**	3.510	[6145, 6159]	6527**	8.472	[6511, 6544]	5583**	15.65	[5552, 5614]
Post-[c]									
Female	5776**	5.281	[5765, 5786]	6368**	11.61	[6346, 6391]	5490**	23.19	[5444, 5535]
Male	5786**	4.377	[5777, 5795]	6395**	9.705	[6346, 6384]	5481**	19.84	[5442, 5520]
Other[b]	5776**	5.230	[5765, 5786]	6369**	11.47	[6347, 6392]	5490**	23.06	[5445, 5535]
Slope									
Pre-[d]									
Female	-58.14**	0.555	[-59.22, -57.05]	-39.65**	1.422	[-42.43, -36.86]	-57.03**	2.483	[-61.90, -52.17]
Male	-71.61**	0.879	[-77.33, -69.88]	-46.33**	2.274	[-50.78, -41.87]	-75.80**	4.700	[-85.01, -66.58]
Other[b]	-63.45**	4.006	[-71.31, -55.60]	-48.05**	8.115	[-63.96, -32.15]	-55.08*	20.42	[-95.11, -15.05]

Table 19 (continued).

Parameter	Ontario			British Columbia			Newfoundland and Labrador		
	$\hat{B}$	*SE*	95% CI	$\hat{B}$	*SE*	95% CI	$\hat{B}$	*SE*	95% CI
Slope									
Post-[e]									
Female	11.69**	0.988	[9.75, 13.62]	-5.893*	2.293	[-10.39, -1.40]	3.571	3.955	[-4.18, 11.32]
Male	1.507	1.505	[-1.44, 4.46]	-2.037	3.635	[-9.16, 5.09]	-17.20*	7.555	[-32.01, -2.39]
Other[b]	14.78*	6.693	[1.66, 27.90]	10.63	13.74	[-16.31, 37.56]	-11.72	31.04	[-72.55, 49.11]

Note. $\hat{B}$ = unstandardized regression coefficient; *SE* = robust standard error; CI = confidence interval.

[a] = study week 12; [b] = identified gender not female or male; [c] = study week 21; [d] = study weeks 1 - 12; [e] = study weeks 18 - 25.

* = $p < .05$; ** = $p < .01$.

Table 20. Estimated weekly mean daily step count intercepts and slopes by gender (between provinces), complete cases sample.

Parameter	**Ontario and British Columbia**			**Ontario and Newfoundland and Labrador**			**British Columbia and Newfoundland and Labrador**		
	$\hat{B}$	***SE***	**95% CI**	$\hat{B}$	***SE***	**95% CI**	$\hat{B}$	***SE***	**95% CI**
Intercept									
Pre-[a]									
Female	-373.3**	9.275	[-394.5, -358.1]	568.9**	16.11	[537.3, 600.4]	945.2**	17.91	[910.1, 980.3]
Male	-382.2**	7.945	[-397.8, -366.7]	579.1**	14.39	[550.9, 607.4]	961.4**	15.86	[930.3, 992.5]
Other[b]	-375.6**	9.170	[-393.6, -357.6]	568.8**	16.04	[537.3, 600.2]	944.4**	17.80	[909.5, 979.3]
Post-[c]									
Female	-592.8**	12.76	[-617.8, -567.8]	285.8**	23.78	[239.2, 332.4]	878.6**	25.93	[827.8, 929.5]
Male	-578.9**	10.65	[-599.7, -558.0]	305.2**	20.32	[265.4, 345.1]	884.1**	22.09	[840.8, 927.4]
Other[b]	-593.7**	12.61	[-618.4, -569.0]	285.8**	23.65	[239.5, 332.2]	879.5**	25.76	[829.0, 930.0]
Slope									
Pre-[d]									
Female	-18.49**	1.526	[-21.48, 15.50]	-1.102	2.544	[-6.09, 3.89]	17.39**	2.861	[11.78, 22.99]
Male	-25.28**	2.438	[-30.06, -20.50]	4.191	4.781	[-5.18, 13.56]	29.47**	5.221	[19.24, 39.70]
Other[b]	-15.40	9.050	[-33.14, 2.34]	-8.375	20.81	[-49.17, 32.42]	7.025	21.98	[-36.05, 50.10]

Table 20 (continued).

Parameter	Ontario and British Columbia			Ontario and Newfoundland and Labrador			British Columbia and Newfoundland and Labrador		
	$\hat{B}$	*SE*	95% CI	$\hat{B}$	*SE*	95% CI	$\hat{B}$	*SE*	95% CI
Slope									
Post-[e]									
Female	17.58**	2.497	[12.69, 22.47]	8.116*	4.077	[0.13, 16.11]	-9.464*	4.572	[-18.42, -0.50]
Male	3.544	3.934	[-4.17, 11.26]	18.71*	7.703	[3.61, 33.81]	15.16	8.384	[-1.27, 31.60]
Other[b]	4.155	15.29	[-25.80, 34.11]	26.50	31.75	[-35.73, 88.73]	22.35	33.94	[-44.18, 88.87]

Note. $\hat{B}$ = unstandardized regression coefficient; *SE* = robust standard error; CI = confidence interval.

[a] = study week 12; [b] = identified gender not female or male; [c] = study week 21; [d] = study weeks 1 - 12; [e] = study weeks 18 - 25.

* = $p < .05$; ** = $p < .01$.

Appendix I: Curriculum Vitae

Curriculum Vitae

Name:	Sean Spilsbury
Post-secondary Education and Degrees:	Western University London, Ontario, Canada 2011-2016 B.A. (Honors Double Major) Western University London, Ontario, Canada 2019-2021 (In Progress) M.A.
Honours and Awards:	Queen Elizabeth II Aiming for the Top Scholarship 2011 Western Scholarship of Distinction 2011 Dean's Honor List 2013, 2015, 2016 Julie Polanksi Memorial Award 2015 Ontario Graduate Scholarship 2019
Related Work Experience	**Research Assistant** Dr. Marc Mitchell's Lab (Western University) **Teaching Assistant** Western University 2020-2021

Publications:

Pentland, V., Spilsbury, S., Biswas, A., Mottola, M.F., Paplinskie, S., & Mitchell, M. (2021). Does walking reduce postpartum depressive symptoms? A systematic review and meta-analysis of randomized controlled trials. *Journal of Women's Health*. doi: 10.1089/jwh.2021.0296

Abstract Publication: Hiemstra, M. Spilsbury, S. Mitchell, M. Oh, P. (2020). Can Financial Incentives Promote Exercise Adherence Amongst Cardiac Rehabilitation Graduates? A 24-week Pilot Randomized Controlled Trial. *Medicine & Science in Sports & Exercise*. *52*(7S), 441. doi: 10.1249/01.mss.0000678696.60517.7b

Posted Presentation: Can Financial Incentives Promote Exercise Adherence Amongst Cardiac Rehabilitation Graduates? A 24-week Pilot Randomized Controlled Trial. American College of Sports Medicine’s 67th General Meeting. San Francisco, California on May 26-30, 2020 (Cancelled).