Defence Research and Development Canada
Recherche et développement pour la défense Canada

CAN UNCLASSIFIED

DRDC | RDDC
technologysciencetechnologie

Testing the Efficacy of the Road to Mental Readiness (R2MR) Mental Health Education and Resilience Training Program During Basic Military Qualification

A Pilot Group Randomized Control Trial

Deniz Fikretoglu
DRDC – Toronto Research Centre

Aihua Liu
Douglas Mental Health University
McGill University

Kristen Blackler
DRDC – Toronto Research Centre

Defence Research and Development Canada
Scientific Report
DRDC-RDDC-2018-R158
October 2018

CAN UNCLASSIFIED

Canada

IMPORTANT INFORMATIVE STATEMENTS

This document was reviewed for Controlled Goods by Defence Research and Development Canada (DRDC) using the Schedule to the *Defence Production Act*.

Endorsement statement: This publication has been peer-reviewed and published by the Editorial Office of Defence Research and Development Canada, an agency of the Department of National Defense of Canada. Inquiries can be sent to: Publications.DRDC-RDDC@drdc-rddc.gc.ca.

Abstract

Background: The Road to Mental Readiness (R2MR) program is the standard mental health education and resilience training program in the Canadian Armed Forces (CAF). The overall goal of R2MR training is to improve military performance, psychological health, resilience, and attitudes towards using mental health services. Since 2008, R2MR has been implemented throughout the military career and deployment cycles and thousands of military personnel have received R2MR. Like any other large-scale workplace mental health intervention, R2MR has to be tested for efficacy to see if it is achieving its program objectives. DRDC – Toronto Research Centre has been asked to conduct a Group Randomized Control Trial (GRCT) to test the efficacy of R2MR during military members' first exposure to the program, at Basic Military Qualification (BMQ).

Objective: A small pilot study was conducted in preparation for the larger GRCT, between October 31st, 2016 and February 8th, 2017. The primary objective of the pilot study was to assess the feasibility of the larger GRCT. The feasibility findings were summarized in an earlier report. The objectives of the current report are to provide descriptive and efficacy findings from the pilot study.

Methods: Eight Anglophone platoons were recruited for the study and randomized to an Intervention or a Delayed Intervention / Control condition. Three data collection sessions took place in Weeks 2, 5, and 9 of the BMQ (Baseline (T1), Follow-up 1 (T2) and Follow-up 2 (T3), respectively). At each data collection session, participants completed questionnaires assessing their psychological health and resilience as well as their attitudes and intentions towards mental health service use. Platoons randomized to the Intervention condition received R2MR in Week 2 of their BMQ (after the first data collection session); those randomized to the Control condition received R2MR in Week 9 of their BMQ (after the last data collection session). Performance outcomes were obtained for those participants who consented to data linkage to an administrative database at the Canadian Forces Leadership and Recruit School (CFLRS); data linkage was performed after the pilot study ended. Mixed effect models were obtained to examine efficacy. For continuous outcomes, we employed mixed linear models assuming random intercepts and slopes to account for platoon-level differences. For binary outcomes, we used generalized linear mixed models to assess individual-level differences while taking into account the platoon-level covariance.

Results: Out of a possible 427 Non-Commissioned Member (NCM) recruits, a total of 354 (82.90%) consented to participate in the study and completed T1 data collection. Of those original 354 participants, 296 completed T2 data collection (83.62%) and 278 completed T3 data collection (78.53%). A total of 267 participants (66.3%) provided consent to data linkage. There were no statistically significant differences between the two conditions on psychological health or resilience at Follow-up 1 or Follow-up 2. For some but not all of the performance outcomes, there was a trend toward beneficial effects. For attitudes and intentions towards mental health service use, there were consistent and statistically significant beneficial effects at Follow-up 2.

Conclusions: In this small pilot GRCT, we found mixed support for the presumed beneficial effects of R2MR. These results must be interpreted with great caution in the context of the well-recognized limitations of small pilot studies (e.g., the risk for Type I and II error) in general, and the specific limitations of this pilot study in particular.

Significance to Defence and Security

Establishing the efficacy of the Road to Mental Readiness (R2MR) mental health education and resilience training program is critical to ensuring CAF members remain resilient in the course of their military careers. This Scientific Report captures the first set of efficacy findings on the R2MR from a pilot group randomized control study.

Résumé

Contexte: En route vers la préparation mentale (RVSM) est le programme de formation standard des Forces armées canadiennes (FAC) axé sur l'éducation en santé mentale et la résilience. L'objectif général du programme RVSM est d'améliorer le rendement, la santé mentale, la résilience et les attitudes à l'égard de l'utilisation des services de santé mentale par les militaires. Depuis 2008, le programme RVSM est mis en œuvre tout au long de la carrière militaire et des cycles de déploiement, et des milliers de militaires y ont déjà pris part. Comme dans toute autre intervention à grande échelle en santé mentale au travail, on a dû tester l'efficacité du programme RVSM afin de s'assurer qu'il atteignait bien ses objectifs. RDDC Toronto s'est vu confier le mandat de procéder à un essai contrôlé randomisé (ECR) par grappes pour évaluer l'efficacité du programme RVSM, lorsque les militaires y prennent part pour la première fois, durant la qualification militaire de base (QMB).

Objectif: Entre le 31 octobre 2016 et le 8 février 2017, on a réalisé une petite étude pilote afin de préparer la tenue de l'ECR par grappes à plus grande échelle. L'objectif premier de l'étude pilote était d'évaluer la faisabilité d'un ECR par grappes de plus grande envergure. Les conclusions quant à la faisabilité sont résumées dans un rapport précédent. Le présent rapport a pour objectif de fournir une description et les résultats quant à l'efficacité tirés de l'étude pilote.

Méthodes: Huit pelotons anglophones ont été recrutés pour l'étude et répartis aléatoirement soit dans le groupe expérimental, soit dans le groupe témoin (avec intervention différée). Trois séances de collecte de données ont eu lieu au cours des semaines 2, 5 et 9 de la QMB (référence [T1], suivi 1 [T2] et suivi 2 [T3], respectivement). Lors de chacune des séances de collecte de données, les participants ont rempli un questionnaire visant à évaluer leur santé mentale et leur résilience, ainsi que leurs attitudes et intentions à l'égard de l'utilisation des services de santé mentale. Les pelotons répartis aléatoirement du groupe expérimental ont pris part au programme RVSM durant la semaine 2 de leur QMB (après la première séance de collecte de données); ceux du groupe témoin ont pris part au programme RVSM durant la semaine 9 de leur QMB (après la dernière séance de collecte de données). On a obtenu des résultats de rendement chez les participants qui avaient consenti au couplage de ces données avec celles d'une base de données administrative de l'École de leadership et de recrues des Forces canadiennes (ELRFC). On a procédé au couplage des données une fois l'étude pilote terminée. Des modèles à effets mixtes ont été obtenus aux fins d'examen de l'efficacité. Dans le cas des résultats continus, nous avons eu recours à des modèles linéaires à effets mixtes avec ordonnée à l'origine et pentes aléatoires pour tenir compte des différences à l'échelle du peloton. Dans le cas des résultats binaires, nous avons utilisé des modèles linéaires généralisés mixtes pour évaluer les différences à l'échelle individuelle tout en tenant compte de la covariance à l'échelle du peloton.

Résultats: Sur 427 participants potentiels parmi les recrues militaires du rang (MR), 354 (82,90 %) ont accepté de prendre part à l'étude et terminé la collecte de données T1. Parmi les 354 participants du début, 296 ont terminé la collecte de données T2 (83,62 %) et 278 la collecte de données T3 (78,53 %). En tout, 267 participants (66,3 %) ont donné leur consentement au couplage des données. Au suivi 1 comme au suivi 2, on n'a noté aucune différence statistiquement significative entre les deux groupes sur le plan de la santé mentale ou de la résilience. Pour certains résultats de rendement, mais pas tous, une tendance vers des effets bénéfiques se dégage. Dans le cas des attitudes et des intentions à l'égard de l'utilisation des services de santé mentale, des effets bénéfiques ressortent de façon constante et statistiquement significative au suivi 2.

Conclusions: Au cours de la petite étude pilote prenant la forme d'un ECR par grappes, nous avons obtenu des résultats mitigés à l'appui des effets bénéfiques présumés du programme RVSM. Ces résultats doivent être interprétés avec la plus grande prudence compte tenu des limites reconnues des petites études pilotes (p. ex. risque d'erreur de première ou de deuxième espèce) en général, et des limites propres à cette étude pilote en particulier.

Importance pour la défense et la sécurité

Il est crucial d'établir l'efficacité du programme de formation RVSM axé sur l'éducation en santé mentale et la résilience afin de s'assurer que les membres des FAC demeurent résilients tout au long de leur carrière militaire. Le présent rapport scientifique fait état de la première série de conclusions sur le rendement du programme RVSM obtenus dans le cadre d'un ECR par grappes.

Table of contents

List of Figures

List of Tables

This page intentionally left blank.

1 Introduction

1.1 Background

The Road to Mental Readiness (R2MR) program is the standard mental health education and resilience training program in the Canadian Armed Forces (CAF). R2MR was developed at the request of the Chief of Military Personnel (CMP) and the CAF Surgeon General, and was launched in 2007 [1]. According to the Director General Health Services (DGHS), "R2MR training encompasses the entire package of resilience and mental health training that is embedded throughout Canadian Armed Forces (CAF) members' career, including the deployment cycle. R2MR training is layered and tailored to meet the relevant demands and responsibilities CAF personnel encounter at each stage of their career and while on deployment. In this way R2MR is designed to ensure that the most appropriate training is provided when required to ensure CAF personnel are prepared mentally for the challenges they may encounter." [2] The first exposure to R2MR for most military personnel takes place soon after entry into the CAF, during Basic Military Qualification (BMQ).

The goals of the R2MR program are to improve psychological health, psychological resilience, and military performance. As well, R2MR aims to remove barriers to care and to encourage individuals to seek mental healthcare if and when needed. A set of core components or learning objectives are included in all versions of R2MR to achieve these goals. These key components or learning objectives are 1) increasing mental health literacy; 2) teaching stress management skills; and 3) changing beliefs, attitudes, and intentions towards mental health service use.

The key learning objectives of R2MR and assumptions regarding its potential beneficial effects are similar to those of other large-scale mental health education and training interventions in various military populations. A small number of randomized control studies have been conducted in the past five years in the United States (U.S.) and United Kingdom (U.K.) to examine the efficacy of these similar interventions. However, the results have been mixed regarding beneficial effects. Where beneficial effects were detected, they did not necessarily extend to all outcomes of interest, were not necessarily maintained over time from immediately post-intervention to short-term follow-up (e.g., 3–6 months), and effect sizes were generally small (.05 to .30 range) [3]–[7]. Smaller-scale, less rigorous (e.g., quasi-experimental) studies in military populations in the U.S. and Australia [8]–[10] of interventions similar to R2MR have also yielded similarly mixed results.

Given i) the various differences between the military populations studied to date (U.S., U.K., and Australia) and the CAF, which limit the generalizability of existing findings to the CAF, ii) the mixed findings in the literature on military mental health education and resilience training interventions, and finally, iii) emerging literature that shows civilian and military workplace mental health education interventions having only small beneficial effects that seem to diminish over time [11], R2MR needs to be tested for efficacy. An efficacy trial can determine if (and to what extent) meaningful changes in the outcomes of interest are indeed taking place, and whether they are maintained over time from immediate post-intervention to longer-term.

While any of the existing R2MR versions could be tested for efficacy, a number of considerations favour choosing the BMQ version: first, the BMQ is military members' first exposure to R2MR and as such provides the foundation upon which all further mental health training is built. Therefore, ensuring that

R2MR is efficacious at BMQ is critical for the success of all mental health training in the CAF. Second, BMQ is the only setting in which there is a captive audience/research participant pool which makes an efficacy study feasible. Third, given the large number of Non-Commissioned Member (NCM) recruits who go through BMQ training on a continuous basis, the BMQ setting provides the largest sample size possible to detect what are likely to be small-size effects [12]. And finally, fourth, a significant portion of existing research on the efficacy of mental health education and training has been conducted in the recruit population / basic military training context; this makes it easier to compare results from an efficacy study in the CAF with those in the existing literature.

1.2 Establishing Efficacy With Group Randomized Control Trials (GRCTs)

Randomized Control Trials (RCTs) are the gold standard for establishing efficacy for a variety of interventions, including medical and/or mental health interventions such as R2MR. In the simplest type of RCT design, participants/individuals are randomly assigned to either an intervention or a control condition. In settings where pre-existing clustering or grouping of individuals is present, where the intervention is delivered at the group (not the individual) level, and where there is "the risk of contamination"—whereby group members randomized to the intervention condition could influence those randomized to the control condition through sharing the active ingredients of the intervention—it is more appropriate to randomize subjects at the group level, i.e., to conduct a group randomized control trial (GRCT) [13]–[16]. In the case of the BMQ, individual recruits go through their BMQ training within a platoon (i.e., there is a pre-existing grouping or clustering of intervention targets), R2MR is delivered at the platoon (i.e., group) level, and the risk of contamination within a platoon (i.e., the group) cannot be ruled out. As such, testing the efficacy of R2MR requires a GRCT.

1.3 The Need for Pilot and Feasibility Studies in GRCTs

A pilot study can be defined as "a version of the main study that is run in miniature to test whether the components of the main study can all work together" [17] (p. 5). Pilot studies are especially important to conduct prior to GRCTs because these designs are very complex and resource-intensive. Furthermore, the operational and training setting in which R2MR is delivered at BMQ poses many challenges to study design [18]. Prior to commencing the full GRCT, there was thus a need to conduct a pilot study to assess the feasibility of the study design and to refine study procedures and methods as needed. This pilot study was conceived to be external to the main study (data collected would not be used in the analysis for the main GRCT). External pilot studies allow for the design of the main study to be changed if necessary based on the findings of the pilot [15].

1.4 Primary GRCT Objectives

Based on the existing literature [3]–[7] and pilot work conducted on R2MR among CAF NCM recruits [19]–[28], we hypothesized for the larger GRCT that:

R2MR will have a beneficial effect on psychological health

R2MR will have a beneficial effect on psychological resilience

R2MR will have a beneficial effect on attitudes towards and intentions for mental health service use

R2MR will have a beneficial effect on performance in BMQ training

R2MR's beneficial effects will be in the very small-to-medium range

R2MR's beneficial effects will diminish over time, from immediately post-intervention to short-term follow-up

1.5 Pilot Study Objectives

The key objectives of the pilot study included:

a. Refining the randomization and scheduling procedures to ensure the intervention and data collection sessions were delivered at the appropriate times

b. Refining the procedures required to ensure the participants, platoon instructors, data collection staff, and the principal investigator all remained blind to participant condition

c. Ensuring the necessary contracting mechanisms were established for data collection staff

d. Refining the data collection procedures and materials

e. Assessing participation rates, attrition, and data validity

f. Refining the procedures for data management

An additional objective of the pilot study was to:

g. Summarize descriptive and efficacy findings

In a separate report, we summarized the pilot feasibility findings regarding randomization, scheduling, blinding, as well as participation, attrition, and threats to data validity (objectives a–f above). The current report summarizes the pilot efficacy findings. In structuring the current report, we followed the Consolidated Standards of Reporting Trials (CONSORT) guidelines for the reporting of pilot GRCTs [29] (See Annex A). Annex B summarizes where each CONSORT requirement can be found in this Scientific Report. Portions of the Methods in the current report (Design, Participants and Procedures) are drawn from the Methods section of the previous report focusing on feasibility [18].

2 Methods

2.1 Design

When we designed the GRCT, all incoming recruit platoons were receiving R2MR at Week 2. Because R2MR had already become part of standard training, it was not possible to have a traditional, "pure," control group that received no R2MR. Instead, a "Delayed Intervention" group that received R2MR close to the end of the BMQ, at Week 9, serves as the Control group.

In the pilot study, participating platoons were randomly assigned to either the Intervention (R2MR at Week 2 of the BMQ) or the Control (R2MR at Week 9 of the BMQ) group. For both the Intervention and the Control groups, three assessments were conducted: prior to R2MR exposure, around Week 2 (a day or two before R2MR: Baseline or T1), around Week 5 (approximately three weeks after exposure to R2MR for the Intervention group: Follow-up 1 or T2), and towards the end of the BMQ around Week 9 (a day or two before the control group receives R2MR: Follow-up 2 or T3). See Figure 1 for study design.

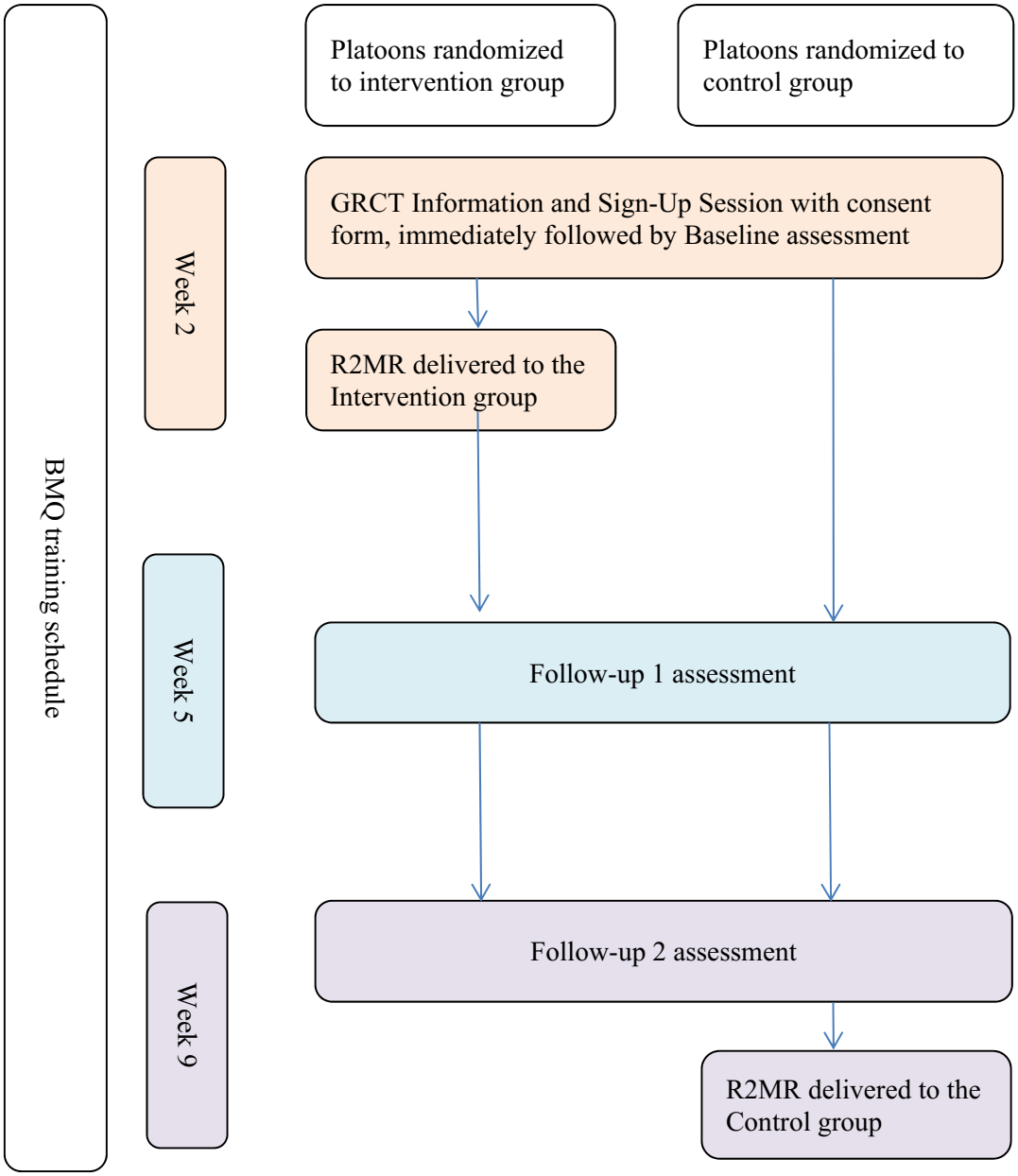


Figure 1: *Overview of the design of the GRCT.*

2.2 Participant Selection

The target participant population for the larger GRCT is all new Anglophone recruit platoons arriving at the Canadian Forces Leadership and Recruit School (CFLRS) for their BMQ training. In any given calendar year, 75% or more of the incoming NCM platoons are Anglophone. (Francophone platoons are excluded from the larger GRCT (please see the study protocol for full justification [16]. To help provide some evidence of the effectiveness of R2MR for Francophone platoons at BMQ, there will be a non-randomized, uncontrolled parallel effectiveness study among Francophone platoons while the GRCT among the Anglophone platoons is running).

Platoons included in the pilot were selected first based on availability. In order to minimize respondent burden and to maximize participation rates, it was important for us not to overlap with another large research study. Platoons became available for our pilot after a multi-year longitudinal study on retention (i.e., Project Horizon) completed data collection towards the end of October 2016.

Platoons included in the pilot were selected also based on their eligibility for the larger GRCT. Platoons that start their BMQ training in October–November interrupt their BMQ training and go home for three weeks for the winter holidays. All other BMQ platoons complete their training without interruption. The platoons that start their BMQ training in October–November thus differ from all other platoons in the CFLRS training calendar in ways that can potentially influence study outcomes. These platoons are therefore ineligible for the larger GRCT and ideal to include in a pilot.

The pilot study described in this Scientific Report included eight Anglophone platoons who started their BMQ training in October and November 2016. (Descriptive findings for the two Francophone platoons from which we collected data in a parallel non-randomized, uncontrolled pilot will be described in a separate report).

All individuals in the selected platoons were eligible to participate and there were no exclusion criteria.

Each platoon in the pilot study completed data collection as a group in a classroom at CFLRS in Saint-Jean-Sur-Richelieu, Quebec. In the initial Week 2 Baseline data collection session, the data collection staff explained the purpose of the study and provided an Information Sheet and an Informed Consent Form. Participants were also given a letter which explained that the study had been endorsed by the Commanding Officer at CFLRS. Participants were then given a chance to ask any questions they had before signing the Informed Consent Form. The Informed Consent Form asked for separate consent for 1) participating in the current study; and 2) providing permission to link study data to recruit research and administrative databases.

2.3 Intervention

As stated previously, R2MR at BMQ has three learning objectives: 1) to increase mental health literacy; 2) to teach stress management skills; and 3) to change beliefs, attitudes and intentions towards mental health service use.

R2MR uses a colour-coded (green, yellow, orange, red) figure, the Mental Health Continuum Model (MHCM), to increase mental health literacy; a bidirectional arrow in the MHCM captures movement along the continuum, indicating that there is always the possibility for a return to full health and functioning; behavioural indicators under each colour category in the MHCM familiarize recruits with basic mental health and mental illness concepts.

To teach stress management skills, R2MR introduces four skills (i.e., the Big 4) to participants: tactical (diaphragmatic) breathing, goal setting, visualization, and self-talk. Self-talk includes both positive mantras (repeating positive thoughts such as "I can do this") and cognitive restructuring. After each skill is defined, the relevance of the Big 4 skills to successful performance in BMQ is explicitly addressed and recruits are given BMQ-specific exercises to help practice the skills.

Following the Big 4 skills, recruits learn how to recognize need for treatment using the MHCM, They are given information about what happens in treatment, and are provided with a list of resources available to individuals who might fall under each of the colour categories in the MHCM. They are also presented with common attitudinal barriers to seeking treatment and provided with ways to overcome these barriers. After these didactic modules, recruits are broken into smaller groups, and are given hypothetical vignettes to help further reinforce their mental health literacy and stress management skills.

R2MR in its entirety is delivered as a 160-minute PowerPoint classroom presentation at BMQ. The main difference between the Week 2 (Intervention) and the Week 9 (Control) versions of the R2MR materials was in the speaker notes for the instructor delivering the material. The Week 2 version emphasized the relevance of the concepts and skills to BMQ training and the Week 9 version, being close to the end of the BMQ, emphasized the relevance of the concepts and skills to later military training.

R2MR sessions (both Intervention and Control) in the pilot phase of this GRCT were delivered by one bilingual instructor, a peer educator (i.e., former military member) who completed the standard R2MR training by DGHS, as well as approximately 20 hours of additional training focusing on intervention adherence (i.e., treatment fidelity) with the principal investigator (D.F.). The same instructor is continuing as the instructor for the GRCT study phase.

2.4 Key Outcomes and Other Measures

R2MR has multiple mental health education and training objectives. Data were collected to assess four main areas:

i. Psychological health

ii. Psychological resilience

iii. Mental Health Service Use (MHSU) beliefs, attitudes and intentions

iv. BMQ performance and graduation rate

In addition, the study assessed several covariates. A covariate is a variable that is possibly predictive of the outcome under study. A covariate may act as a confounder/mediator/moderator of the effects of R2MR. Covariates included:

v. Cognitive functioning/intelligence

vi. Personality

vii. Social desirability

viii. Mental health literacy

ix. Stress management skills

x. Sociodemographic and military variables

These measures are described in greater detail below.

2.4.1 Outcome Measures

2.4.1.1 Psychological Health

2.4.1.1.1 Patient Health Questionnaire (PHQ-9)

Severity of current depressive symptoms (past two weeks) was assessed by the Patient Health Questionnaire (PHQ) [30]. The PHQ is a brief, single factor, 9-item self-report questionnaire with well-established reliability, validity, and sensitivity [31]–[33]. PHQ depression severity cut-offs are as follows: mild (5–9), moderate (10–14), moderately severe (15–19), severe (> 19) [34].

2.4.1.1.2 K-10 and Subjective Units of Distress Scale (SUDS)

Psychological distress was assessed using two self-report measures, the Kessler Psychological Distress Scale (K-10) [35] and the Subjective Units of Distress Scale (SUDS) [36]. The K-10 is a 10-item questionnaire assessing items such as nervousness, agitation, fatigue, and negative affect. Good internal consistency (α = .89 to .92) and construct validity of the K-10 have been established in the general population and military samples [35], [37], [38]. The following K-10 bands are suggested: low (10–15), moderate (16–21), high (22–29), and very high (30–50) [39]. In a sample (N = 1,264) of deployed CAF personnel, a cut-off of 16 points on the K10 classified self-reported occupational impairment with a sensitivity of 88% and specificity of 75% [37]. The SUDS [36] is a one-item self-report that provides an estimate of current severity of subjective distress, anxiety, fear or discomfort on a scale from 0 to 100. Previous studies utilizing the SUDS have shown preliminary evidence of satisfactory concurrent validity [36], [40].

2.4.1.1.3 GAD-7

The Generalized Anxiety Disorder Scale (GAD-7) [41] was used as a measure of anxiety. The GAD-7 is a 1-factor, 7-item, self-report questionnaire that has demonstrated good internal consistency ($\alpha = .89$) and validity in both the general population and primary care samples [41]–[43]. GAD-7 scores can be categorized into mild (5–9), moderate (10–14), and severe (15–21) levels of anxiety [42]. A cut-off of 10 points classifies the presence of generalized anxiety disorder with a sensitivity of 89% and specificity of 82% [44].

2.4.1.2 Psychological Resilience

2.4.1.2.1 CD-RISC

Psychological resilience was assessed using the Connor-Davidson Resilience Scale (CD-RISC), in its original version a 25-item self-report questionnaire developed by K.M. Connor and J.R. Davidson [45]. The CD-RISC has been widely used in community, clinical, and military samples and has demonstrated good internal consistency and construct validity for the original version [45], as well as a 10-item abbreviated version [46]. We used the 10-item abbreviated version in order to minimize respondent burden.

2.4.1.3 MHSU Attitudes and Intentions

2.4.1.3.1 CAF-MHSUQ

MHSU attitudes and intentions were assessed with the Canadian Armed Forces Mental Health Service Use Questionnaire (CAF-MHSUQ) [47], a 90-item self-report measure developed specifically to assess MHSU intentions among CAF recruits. In addition to assessing MHSU intentions, CAF-MHSUQ assesses MHSU (affective and instrumental) attitudes, MHSU subjective norms (i.e., perceptions of how supportive important others would be of MHSU), and MHSU perceived behavioural control (i.e., perceptions of how much control individuals believe they have over MHSU, how difficult they perceive MHSU to be, and how confident they feel about overcoming barriers to MHSU). CAF-MHSUQ also assesses beliefs, expectations, and thoughts that may be driving overall MHSU attitudes, subjective norms, perceived behavioural control, and intentions. CAF-MHSUQ was developed based on the widely used Theory of Planned Behavior [48] and has shown good psychometric properties (i.e., internal consistency reliability estimates and factorial/structural validity) in multiple studies [49], [50]. Because the CAF-MHSUQ is a measure under development, psychometric analyses (results to be published in a separate report) were conducted with the pilot data in order to determine optimal scoring for each of the subscales.

2.4.1.4 BMQ/Military Performance

An important outcome for R2MR is military performance, in this case, performance during BMQ training. Historically, graduation rates have varied across different platoons, with as much as 25% of incoming NCM recruits not graduating in some years. CFLRS has an administrative database which includes BMQ graduation information. CFLRS also has additional administrative databases that include

performance measures (e.g., results of the Fitness for Operational Requirements of CAF Employment [i.e., FORCE] test, the First Aid test, the Weapons Handling test (Pass/Fail), The Weapons Shooting test (out of a possible 25 points maximum, with minimum 15 point required for passing), and the 13-K March). We used graduation status as our primary military performance outcome; based on discussions with CFLRS staff, we also explored additional performance outcomes such as the results from the FORCE test at Week 8, the Weapons Handling and Shooting Tests (Pass/Fail status and continuous score), the First Aid test score, and Pass/Fail status on the 13-K March. We obtained BMQ performance information through a data linkage with CFLRS recruit administrative databases. We asked participants for permission to link their data from the CFLRS administrative databases with the current study. On the Informed Consent Forms, this consent for permission of data linkage was separate from the consent to participate in the current study.

2.4.1.5 Covariates

2.4.1.5.1 Sociodemographic Information

A Sociodemographic Questionnaire developed specifically for this study was used to assess age, gender, ethnicity, education, and self-reported physical and mental health, all possible moderators of the R2MR intervention effect.

2.4.1.5.2 Cognitive Aptitude (Shipley-2)

In previous research among NCM recruits, cognitive aptitude was found to have medium-to-large effects on intermediate learning outcomes in R2MR [18], [20], [26]. We therefore included cognitive aptitude as a potential confounder. Cognitive aptitude was assessed by the Shipley-2 [51], a 30-minute three-part intelligence test measuring performance on verbal (crystalized intelligence), abstraction (fluid intelligence), and block pattern recognition (fluid intelligence). The Shipley-2 has acceptable levels of internal reliability (split half .91) and test-retest reliability (correlation range from .74–.94). Concurrent validity has been demonstrated between the Shipley-2 and various measures of intelligence. The Shipley-2 can produce a single Intelligence Quotient (IQ) score by combining the verbal subscale and one of the fluid intelligence subscales. Based on analyses conducted as part of prior R2MR research in this population (available upon request), we determined that it would be sufficient to include only the block pattern recognition task using the Block Design subscale of the Shipley to control for the potential confounder of cognitive aptitude. This minimizes respondent burden, and provides a measure not influenced by language.

2.4.1.5.3 Marlowe-Crowne Social Desirability (MC-SDS)

Social desirability can influence responses to self-report questionnaires; our pilot work in the CAF NCM recruit population suggests that social desirability may influence reporting on questionnaires assessing personality, psychological health and functioning, and attitudes and intentions. We therefore examined and controlled for the effects of social desirability. The 33-item Marlowe-Crowne Social Desirability Scale (MC-SDS [52]) was designed to measure social desirability independent of psychopathology. MC-SDS assesses whether respondents are responding truthfully or are misrepresenting themselves. Psychometric studies of the MC-SDS have identified a number of possible scoring schemes and previous psychometric studies in the target population of NCM recruits found only partial support for some of the

existing scoring schemes [53]. We therefore conducted psychometric analyses (results to be published in a separate report) with the pilot data in order to determine optimal scoring.

2.4.1.5.4 Stress Management Skills Test of Performance Strategies (TOPS)

Five subscales, each consisting of four items, of the Test of Performance Strategies (TOPS [54]) was used to assess self-reported use of stress management skills taught in R2MR: goal setting, self-talk, control of negative thinking, mental imagery (i.e., visualization), and relaxation. Similar to previous research in military settings [4], referents were adapted for a military context. The TOPS subscales have demonstrated good internal consistency reliabilities, with Cronbach's alphas of .82 to .93 (goal setting), .76 to .83 (self-talk), .80 to .93 (imagery), .75 to .83 (negative thinking) and .70 to .84 (relaxation) [4], [55]. These subscales have shown convergent validity (goal setting, imagery, relaxation and self-talk [55]) and divergent validity (negative thinking [55]) with some or all subscales of the Athlete Engagement Questionnaire (Confidence, Dedication, Vigor, Enthusiasm). Additionally, negative thinking has been found to correlate positively with all subscales of the Athlete Burnout Questionnaire [55]. The evidence for the structural validity of the TOPS, however, has been mixed and the TOPS has continued to evolve over the years with the addition of new items and removal of poorly constructed items. We therefore conducted psychometric analyses with the pilot data (results to be published separately) to determine optimal scoring.

2.4.1.5.5 Mental Health Literacy (R2MR Program Evaluation Form)

R2MR training uses the Mental Health Continuum Model to teach basic mental health literacy skills (i.e., understanding mental health and mental illness, understanding mental health lies on a continuum, recognizing signs and symptoms at various levels of health and illness). Importantly, R2MR does NOT teach symptoms of specific mental disorders. Thus, existing measures of mental health literacy, which focus on recognition of the common symptoms of specific disorders (e.g., depression, PTSD) [56], [57] were not appropriate assessment tools for mental health literacy in this study. We therefore used items from a questionnaire, the R2MR Program Evaluation form, developed by the R2MR stakeholders to assess two aspects of mental health literacy, knowledge of basic mental health concepts, and confidence in using available resources to help self and others when mental health issues do arise [58]. Neither the English nor the French version had undergone any psychometric work to date. We therefore conducted psychometric analyses (results to be published in a separate report) with the pilot data to determine optimal scoring.

2.4.1.6 Other Measures

2.4.1.6.1 Intervention Adherence (i.e., Treatment Fidelity)

The intervention condition was assessed for treatment fidelity, with the observer (S.O., of the DGHS) using a systematic Fidelity Checklist to determine whether key R2MR components were covered. The Fidelity checklist has been developed and used in the context of a 4-year program of research on R2MR at BMQ. A quarter of the Intervention sessions were observed for intervention adherence/treatment fidelity. The observer (S.O.) had been trained as an R2MR instructor and had also received additional training from the Principal Investigator of the study (D.F.) on how to use the Fidelity Checklist.

2.5 Pilot Sample Size

Sample size for the pilot was based on the main objective of the pilot study, as well as practical considerations. The main objective for the pilot was to test the feasibility of the study design and procedures. Based on prior research we conducted in the same target population and setting, we expected that 3–4 platoons per study condition would be sufficient to test feasibility. We could have collected data from a larger sample in the pilot phase; however this would have pushed back the start date of the GRCT even more. Considering that the larger GRCT is expected to run for 1.5 years [59], [60], and based on discussions with CFLRS staff, we decided to minimize the duration of the pilot by limiting it to the eight platoons that would normally be ineligible for the larger GRCT.

2.6 Randomization

For the eight Anglophone platoons in the pilot study, blocked randomization was utilized to ensure that four platoons were in the intervention condition and four platoons were in the control condition. Block randomization divides platoons into blocks with the same size ($2n$, where n is an integral), then randomizes platoons in each block so that the same number of platoons (n) are assigned to Group 1 (i.e., intervention) and 2 (i.e., control). Compared to simple randomization, which does not guarantee equal numbers between study arms, blocked randomization has the advantage of ensuring that the number of platoons in Intervention (Group 1) and Control (Group 2) conditions are balanced at any stage of the trial [61], [62]. Randomization was generated by the biostatistician for the GRCT (A.L.) using Random Allocation Software – Version 1.032 [63] and provided to CFLRS several months in advance of the pilot.

2.7 Allocation, Implementation, and Blinding

The GRCT is designed to be triple-blinded, meaning that the principal investigator, the participants, and those administering the data collection sessions will be shielded from information regarding intervention assignment.

For the pilot study, we did not attempt blinding of the principal investigator as we wanted the principal investigator to be able to fully investigate problems with randomization and allocation and to be able to communicate with the CFLRS Standards and Scheduling Divisions to resolve problems. We initially planned for the data collection sessions to be administered by contractors who would be blind to randomization and allocation. Unfortunately, delays in contracting necessitated the use of DRDC Toronto staff to administer the data collection sessions; these staff members were not blind to the randomization scheme generated by the biostatistician.

To achieve blinding and allocation concealment from the pilot study participants, we told pilot study participants only that all NCM recruits would receive R2MR during the BMQ and that the study they would be participating in is intended to examine the efficacy of R2MR by examining psychological health, resilience, attitudes, and performance at three assessment points during the BMQ. We specifically did not tell them that half the platoons would receive R2MR at Week 2 and half at Week 9.

2.8 Funding, Ethics Approval, and Trial Registration

The R2MR GRCT and the pilot study were funded by DGHS. The funders did not play a role in study design, analysis, or interpretation of the pilot findings. The protocol for the study was reviewed and approved by the DRDC Toronto Human Research Ethics Committee, DRDC Toronto Protocol 2016-021.

The R2MR GRCT is registered at: http://www.isrctn.com/ISRCTN52557050.

2.9 Analytic Methods

Some recruits going through BMQ training pause their training due to not meeting fitness requirements, getting injured or falling sick. Once ready to resume their training, these recruits may then be moved into a new platoon that is different than the platoon that they started their BMQ training in (i.e., the recruits get "recourse"). This poses obvious problems for GRCTs as recoursed recruits may move from Intervention to Control groups (and vice versa). For this reason, we decided to remove recoursed recruits from the main efficacy analyses, although we collected data from them. We decided to collect data from recoursed recruits in order to be able to conduct additional analyses to determine whether adding back in recoursed recruits changed efficacy results or not. We used information from the CFLRS administrative database to create a variable indicating whether a given participant was a "new" or a "recoursed" recruit.

Data collected in this pilot study are clustered. Recruits complete their BMQ and R2MR training as groups/platoons and randomization can only be done at the group/platoon level. Recruits within the same platoons share experiences with each other and will become more similar as BMQ training unfolds; thus, recruits within a platoon are more likely to have similar responses and performance metrics than recruits from other platoons. Analysis of such data without considering the clustered nature of the data would result in underestimated variance, possibly leading to spurious findings. To account for the clustered nature of the data, we used mixed effect models to examine efficacy. For continuous outcomes, we employed mixed linear models assuming random intercepts and slopes to account for platoon level differences [64], [65]. For binary outcomes, we used generalized linear mixed models [66] to assess individual-level differences while taking into account the platoon-level covariance.

For outcomes that were included in both Follow-up 1 and Follow-up 2 assessments, we assessed if R2MR improves individual-level outcomes at Follow-up 1 and Follow-up 2 assessments, separately. The following variables were adjusted for in all the models for assessing R2MR effects: baseline outcome (if available), age, gender, ethnicity, education, self-reported physical health status, self-reported mental health status, K-10 score, SUDS score, GAD-7 score, PHQ-9 score, CD-RISC score, the Shipley score, and the MC-SDS score. In calculating the least squares means in the mixed linear models, we used inverse-probability-of-attrition-weights to account for the potential bias due to differential attrition between intervention and control conditions.

In testing the potential beneficial effects of R2MR on each of the main categories of outcomes (i.e., psychological health, mental health service use attitudes and intentions, and military performance) we did not employ multiple correction to adjust the p-values for statistical significance, for two reasons. First, the outcome variables in the same category (e.g., psychological health) are correlated. For example, correlations among the psychological health outcomes of psychological distress, depression, and generalized anxiety were all greater than 0.71. Given that these variables are correlated, it is highly plausible that similar results will be obtained for them from statistical testing. In this kind of situation,

Type I error rate is far less inflated than in the case where all the outcome variables used in statistical testing are independent of each other. In other words, applying multiple correction to adjust p-values for significance testing is inappropriate when the outcomes are correlated as the correction has the potential of leading to under-reporting of study findings [67], [68]. Furthermore, this study is a pilot study where there is low statistical power for detecting significant effects based on our small sample. The pilot study nature also means that we would like to explore possible findings and we pay less attention to statistical testing than the clinical significance of the findings. We agree with others that "it is better to tolerate findings that may later prove to be false than to prematurely discard potentially useful observations because of Type 2 errors caused by corrections for multiplicity" [69].

In the presence of results showing statistically significant differences between the Intervention and Control conditions, we calculated and reported effect sizes to quantify the beneficial effects of R2MR on continuous outcomes. Cohen's d was computed as the difference in the mean scores of two samples divided by the pooled standard deviation [70]. Effect sizes of .2 are considered "small," .5 "medium," and .8 or above "large." For binary outcomes, Odds Ratios (ORs) for the association between R2MR and each of the outcomes were calculated to quantify the effect sizes of R2MR. We conducted all analyses using Statistical Analysis Software (SAS), Version 9.4 (SAS Institute, Cary, North Carolina) [71].

3 Results

3.1 Participation and Attrition Rates

We discussed the details of participation and attrition rates in the first two reports from the pilot, focusing on feasibility and an updated power analysis, respectively [18]. To re-iterate and briefly summarize participation and attrition rates here, "out of a possible 427 potential participants, a total of 354 (82.90%) agreed to participate in the study and completed T1 or baseline data collection. Of those original 354 participants, 296 completed T2 (Follow-up 1) data collection (83.62%) and 278 completed T3 (or Follow-up 2) data collection (78.53%). Two attrition rates were calculated (T1 to T2, and T1 to T3). The overall attrition rate for T1–T2 was 58/354 (16.38%) and the overall attrition rate for T1–T3 was 76/354 (21.47%)" [18] (p. 14). Consent for data linkage to the CFLRS administrative database (and other administrative and research databases) was provided by 267 (66.3%) participants.

3.1.1 Participant Flow

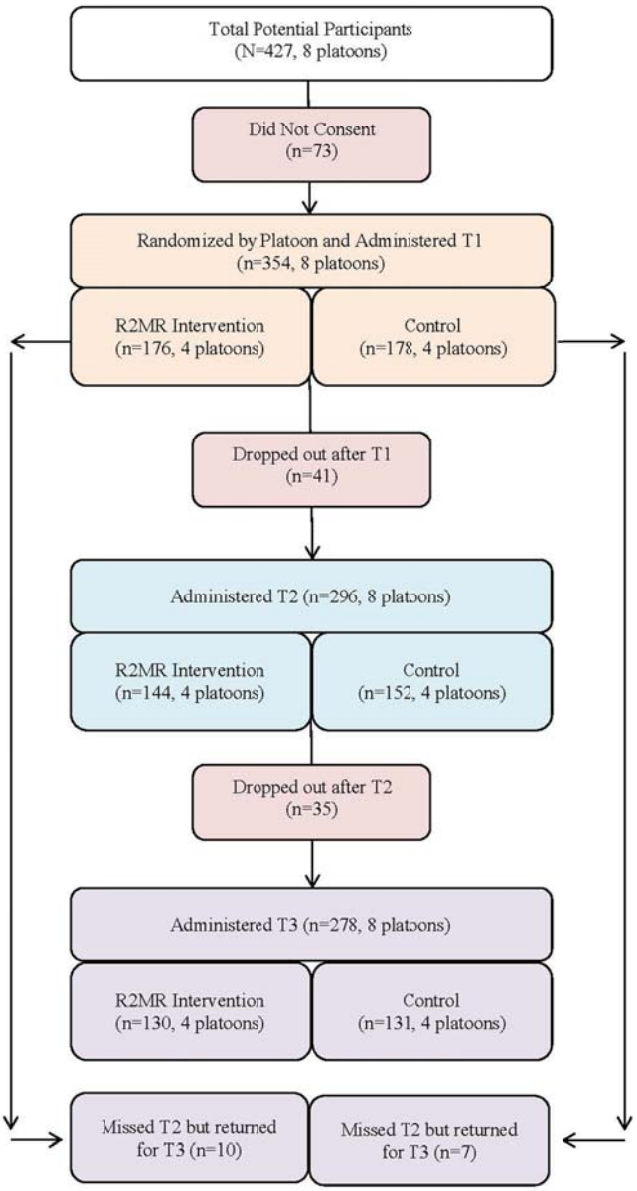


Figure 2: *Participant flow.*

3.2 Recruitment

Participants for the pilot were recruited between October 31, 2016 and February 8, 2017. The pilot ended after data were collected from the eight platoons since the pilot objectives were to test the feasibility of the larger GRCT with only eight platoons.

Table 1: *Pilot dates for Anglophone platoons.*

Platoon Number	BMQ Start Date	Condition	Data Collection Time 1	R2MR Date	Data Collection Time 2	Data Collection Time 3
182	10/31/2016	Control	11/10/2016	01/27/2017	11/30/2016	01/24/2017
183	10/31/2016	Intervention	11/08/2016	11/09/2016	11/29/2016	01/25/2017
186	11/07/2016	Control	11/16/2016	02/02/2017	12/07/2016	01/31/2017
187	11/07/2016	Intervention	11/15/2016	11/18/2016	12/06/2016	02/01/2017
188	11/07/2016	Control	11/17/2016	02/01/2017	12/05/2016	01/30/2017
189	11/14/2016	Intervention	11/22/2016	11/24/2016	01/12/2017	02/06/2017
190	11/14/2016	Intervention	11/23/2016	11/25/2016	01/11/2017	02/08/2017
191	11/14/2016	Control	11/24/2016	02/09/2017	01/12/2017	02/07/2017

3.3 Baseline Data

Baseline characteristics (age, gender, education, ethnicity, self-reported physical and mental health) of the Intervention and the Control platoons were reported in a previous Scientific Report focusing on feasibility [18]. The participants were young, predominantly White and male, with the majority reporting good-to-very good physical and mental health at Baseline. Of note, there were no statistically significant differences on baseline characteristics between the Intervention and Control conditions.

As can be seen from Table 2, recruits in both conditions reported moderate levels of psychological distress on the K-10 and the SUDS, mild symptoms of anxiety and depression on the GAD-7 and the PHQ-9, and moderate levels of perceived psychological resilience at Baseline. This pattern of results corresponds to a relatively psychologically healthy population that may be undergoing some temporary stress (i.e., BMQ training) that may cause short-term psychological distress.

Table 2: *Baseline scores for the main study outcomes.*

Baseline scores	Intervention group		Control group	
	n	*Mean, SD*	*n*	*Mean, SD*
K-10 total score	175	19.05, 7.26	178	20.21, 6.75
SUDS score	175	42.00, 25.44	178	41.04, 23.21
GAD total score	174	7.76, 5.13	178	7.79, 5.61
PHQ-9 total score	173	6.71, 5.66	177	6.35, 4.94
Resilience (CD-RISC) total score	175	28.99, 5.78	178	28.97, 5.03
Attitude (MHSU)				
Instrumental attitude	166	5.54, 1.43	165	5.60, 1.36
Affective attitude	166	4.02, 1.48	164	4.06, 1.44
Intention	173	5.16, 1.40	171	5.21, 1.35
Self-efficacy	168	4.81, 1.19	167	5.06, 1.06
Control	168	5.59, 1.15	168	5.66, 1.14
Subjective norms	172	5.61, 1.06	169	5.53, 1.06
Overall	169	5.16, 0.87	168	5.22, 0.84

3.4 Outcomes

3.4.1 Psychological Health and Psychological Resilience

As can be seen from the Tables 3 and 4 below, there were no statistically significant differences on psychological health and resilience between the Intervention and Control groups at Follow-up 1 or Follow-up 2. Given the absence of statistically significant differences between the two conditions, no effect sizes were calculated. The adjusted mean scores for the Intervention and Control conditions were almost identical at Follow-up 1 and very similar at Follow-up 2 for psychological resilience. The adjusted mean scores for psychological health were slightly, albeit non-significantly, lower for the Control condition at Follow-up 1. At Follow-up 2, the adjusted mean scores for psychological health were either identical or again slightly lower for the Control group. These findings do not support the hypothesis that R2MR has beneficial effects on psychological health or psychological resilience.

Table 3: *T2 (Follow-up 1) Adjusted least squares means from mixed linear model for assessing the effect of R2MR using inverse-probability-of-attrition-weights (Psychological Health and Psychological Resilience).*

T2 outcomes	Intervention group	Control group	Difference (intervention—control)		
			Estimates (95% CI)	Cohen's *d*	*p*-value
K-10 total score	18.60	17.49	1.11 (-0.85–3.06)	-	0.27
SUDS score	30.81	28.45	2.56 (-2.72–7.85)	-	0.34
GAD-7 total score	5.99	5.77	0.21 (-1.46–1.88)	-	0.80
PHQ-9 total score	5.81	5.21	0.59 (-0.37–1.56)	-	0.23
Resilience (CD-RISC) total score	29.73	29.72	-0.01 (-1.34–1.37)	-	0.99

Note: For the K-10, SUDS, GAD-7 and the PHQ-9, higher scores indicate greater psychological distress (or worse psychological health). For the CD-RISC, higher scores indicate greater psychological resilience. The least squares means were calculated with the adjustment for baseline outcome, age, gender, ethnicity, education, self-reported physical health status, self-reported mental health status, K-10 score, SUDS score, GAD-7 score, PHQ-9 score, CD-RISC resilience score, the Shipley score, and the MC social desirability score. In addition, the calculation used inverse-probability-of-attrition-weights to account for the potential bias due to differential attrition.

Table 4: *T3 (Follow-up 2) Adjusted least squares means from mixed linear model for assessing the effect of R2MR using inverse-probability-of-attrition-weights (Psychological Health and Psychological Resilience).*

T3 outcomes	Intervention group	Control group	Difference (intervention—control)		
			Estimates (95% CI)	Cohen's *d*	*p*-value
K-10 total score	17.48	17.48	-0.004 (-1.70–1.69)	-	1.00
SUDS score	27.12	26.89	0.23 (-6.82–7.29)	-	0.95
GAD-7 total score	4.88	4.82	0.06 (-1.09–1.20)	-	0.92
PHQ-9 total score	4.58	4.42	0.16 (-1.09–1.41)	-	0.80
Resilience (CD-RISC) total score	30.69	30.10	0.59 (-1.47–2.65)	-	0.57

Note: For the K-10, SUDS, GAD-7 and the PHQ-9, higher scores indicate greater psychological distress (or worse psychological health). For the CD-RISC, higher scores indicate greater psychological resilience. The least squares means were calculated with the adjustment for baseline outcome, age, gender, ethnicity, education, self-reported physical health status, self-reported mental health status, K-10 score, SUDS score, GAD-7 score, PHQ-9 score, CD-RISC resilience score, the Shipley score, and the MC social desirability score. In addition, the calculation used inverse-probability-of-attrition-weights to account for the potential bias due to differential attrition.

3.4.2 Mental Health Service Use Attitudes and Intentions

There were no statistically significant differences between the Intervention and Control conditions on any of the CAF-MHSUQ subscale scores at the first Follow-up. However, there was a trend towards significance on the Overall Scale score, with the Intervention group reporting more favourable attitudes/intentions compared to the Control condition (Cohen's d = .23). The results were in the same direction (more favourable attitudes and intentions for the Intervention group) but not significant on the other subscales.

By the second Follow-up, the differences between the Intervention and Control groups became statistically significant for the Overall Scale, as well as for all but one of the subscales. The effects sizes ranged from small to medium (Cohen's d = .24 to .37).

Table 5: *T3 (Follow-up 2) Adjusted least squares means from mixed linear model for assessing the effect of R2MR using inverse-probability-of-attrition-weights (Mental Health Service Use Attitudes and Intentions).*

T2 outcomes	Intervention group	Control group	Difference (intervention—control)		
			Estimates (95% CI)	Cohen's*d*	*p*-value
CAF-MHSUQ					
Instrumental attitude	5.75	5.59	0.16 (-0.16–0.47)	-	0.32
Affective attitude	4.08	3.95	0.14 (-0.24–0.51)	-	0.47
Self-efficacy	5.02	4.83	0.19 (-0.14–0.52)	-	0.26
Control	5.78	5.54	0.24 (-0.061–0.55)	-	0.12
Subjective norms	5.74	5.63	0.10 (-0.14–0.35)	-	0.40
Intention	5.17	4.95	0.22 (-0.083–0.52)	-	0.16
Overall Scale	5.25	5.06	0.19 (-0.015–0.40)	0.23	0.068

Note: For the CAF-MHSUQ, higher scores indicate more favourable attitudes and intentions towards mental health service use. The least squares means were calculated with the adjustment for baseline outcome, age, gender, ethnicity, education, self-reported physical health status, self-reported mental health status, K-10 score, SUDS score, GAD-7 score, PHQ-9 score, CD-RISC resilience score, the Shipley score, and the MC social desirability score. In addition, the calculation used inverse-probability-of-attrition-weights to account for the potential bias due to differential attrition.

Table 6: *T3 (Follow-up 2) Adjusted least squares means from mixed linear model for assessing the effect of R2MR using inverse-probability-of-attrition-weights (Mental Health Service Use Attitudes and Intentions).*

T3 outcomes	Intervention group	Control group	Difference (intervention—control)		
			Estimates (95% CI)	Cohen's *d*	*p*-value
Attitude (MHSU)					
Instrumental attitude	5.66	5.54	0.13 (-0.18–0.43)	-	0.41
Affective attitude	4.49	4.13	0.36 (0.052–0.68)	0.25	0.023
Self-efficacy	5.26	4.99	0.27 (0.021–0.52)	0.25	0.034
Control	5.65	5.31	0.34 (-0.014–0.69)	0.30	0.060
Subjective norms	5.67	5.42	0.25 (-0.0034–0.51)	0.24	0.053
Intention	5.27	4.92	0.34 (0.063–0.62)	0.25	0.017
Overall Scale	5.35	5.04	0.31 (0.12–0.50)	0.37	0.002

Note: For the CAF-MHSUQ, higher scores indicate more favourable attitudes and intentions towards mental health service use. The least squares means were calculated with the adjustment for baseline outcome, age, gender, ethnicity, education, self-reported physical health status, self-reported mental health status, K-10 score, SUDS score, GAD-7 score, PHQ-9 score, CD-RISC resilience score, the Shipley score, and the MC social desirability score. In addition, the calculation used inverse-probability-of-attrition-weights to account for the potential bias due to differential attrition.

3.4.3 BMQ Graduation Status and Other Military Performance Outcomes

The BMQ graduation rate was very similar between the Intervention (92.59%) and the Control groups (90.35%). We compared the probability of graduation across the two conditions, controlling for the same baseline measures we controlled for in the mixed models for our continuous outcomes. There was a trend towards a significant difference between the Intervention and the Control groups (Odds ratio = 2.72, 95% CI: 0.76–9.72, *p*-value = 0.13). The results suggest that the probability of graduation was 2.72 times

higher in the Intervention group than the Control group indicating that R2MR may have beneficial effects on BMQ graduation rate. However, as can be seen from the large confidence interval, the effect is of great imprecision/uncertainty.

In addition to BMQ graduation rate, we were able to look at six additional intermediate military performance measures that might be of interest in the BMQ context. The First Aid scores were very similar across the Intervention and Control groups, 95.77 and 95.42, respectively. Results from the mixed linear models controlling for the same baseline measures as in all other mixed models indicated that the difference was not statistically significant (difference = .35, p-value = 0.69). Similarly, the FORCE test scores at Week 8 across the Intervention and the Control groups were very similar, 90.22 and 90.75, respectively. Results from the mixed linear models controlling for the same baseline measures as in all other mixed models plus Week 1 FORCE Test scores indicated that the difference was not statistically significant (difference = -0.53, p-value = 0.40). For the 13-K March success, success rate was about 10% higher in the Intervention than the Control groups, 93.14% and 83.96%, respectively. We compared the probability of 13-K March success across the two conditions, with the same adjustments as those for BMQ graduation rate. There was a trend towards a significant difference between the Intervention and the Control groups (Odds ratio = 5.53, 95% CI: 0.72–42.45, p-value = 0.10). The results suggest that the probability of 13-K March success was 5.53 times higher in the Intervention group than the Control group indicating that R2MR may have beneficial effects on 13-K March success. However, as can be seen from the very large Confidence Interval, the effect is of great imprecision/uncertainty. For the Weapons Handling test, success rate was similar across the Intervention and Control groups, 92.73% and 90.60%, respectively. The difference between the two groups was not statistically significant (Odds ratio = 1.32, 95% CI: 0.43–4.06, *p*-value = 0.63). For the Weapons Shooting test, the scores were similar across the Intervention and the Control groups, 21.93 and 20.90, respectively. The difference was not statistically different (difference = 1.02, se = 0.90, p-value = 0.26). Finally, for the Gas Hut, success rate was higher for the Intervention compared to the Control group (92.73% versus 88.03%, respectively) but this difference was not statistically significant (Odds ratio = 1.73, 95% CI: 0.65–4.56, *p*-value = 0.27).

In addition to looking at group differences on key outcomes, we explored group differences at Time 3 (Follow-up 2) on two measures of interest, the TOPS and the Mental Health Literacy (MHL) scale. The TOPS measures the use of performance strategies that are very similar to and in some cases identical to those stress management skills taught in R2MR (i.e., Diaphragmatic Breathing/Relaxation, Visualization/Imagery, Goal-Setting, and Positive-Negative Self-Talk). The MHL measures knowledge of basic mental health concepts and confidence in using available resources to help self and others when mental health issues do arise. By exploring group differences in these two measures, we wanted to explore the extent to which the stress management and mental health literacy skills taught in R2MR had been taken up by the Intervention group. We wanted to look at the uptake of key learning objectives in R2MR partly based on our previous research in this population and setting showing that many recruits fail to understand and apply the skills taught in R2MR [19], [20], [26], [27]. For the pilot, if R2MR skills were learned and practiced, there should have been large differences between the Intervention and the Control groups by the last assessment point, Time 3 (Follow-up 2). For both the MHL and the TOPS subscales, the Intervention group scores were higher. However, as can be seen in Table 7, on the MHL, the difference between the Intervention and the Control group was relatively small and was not statistically significant. Similarly on the four subscales of the TOPS, the differences across the two conditions were quite small and not statistically significant.

Table 7: *T3 (Follow-up 2) Adjusted least squares means from mixed linear model for assessing the effect of R2MR using inverse-probability-of-attrition-weights (MHL and TOPS).*

T3 outcomes	Intervention group	Control group	Difference (intervention—control)		
			Estimates (95% CI)	Cohen's *d*	*p*-value
MHL	3.94	3.81	0.13	-	0.14
TOPS					
Positive/negative thinking	3.69	3.62	0.07 (-0.12–0.25)	-	0.47
Imaginary	3.51	3.41	0.10 (-0.18–0.37)	-	0.47
Goal setting	3.54	3.36	0.18 (-0.086–0.45)	-	0.18
Relaxation	3.08	3.04	0.04 (-0.20–0.29)	-	0.72

Note: For the MHL, higher scores indicate greater knowledge of basic mental health concepts and greater confidence in using available resources to help self and others when mental health issues do arise. For the TOPS, higher scores indicate more frequent use of the specific stress management skill in question. The least squares means were calculated with the adjustment for baseline outcome, age, gender, ethnicity, education, self-reported physical health status, self-reported mental health status, K-10 score, SUDS score, GAD-7 score, PHQ-9 score, CD-RISC resilience score, the Shipley score, and the MC social desirability score. In addition, the calculation used inverse-probability-of-attrition-weights to account for the potential bias due to differential attrition.

3.4.4 R2MR Intervention Fidelity

The completed Intervention Fidelity Checklists for the two sessions that were observed can be found in Annex C. As can be seen, the instructor adhered perfectly to the standard R2MR material in both sessions and there were no significant omissions or deviations from the standard material, nor any significant insertions of new, contradictory material.

3.5 General Discussion

The main purpose of the pilot GRCT on R2MR was to examine the feasibility of the proposed efficacy trial, using a small, convenience sample. Feasibility in the current pilot was examined by looking at participation and attrition rates, threats to validity, the success of randomization, and blinding, the success of scheduling (intervention and data collection), intervention fidelity, and data collection and management procedures. We summarized the results focusing on feasibility in a separate report [18]; the overall assessment of that report was that the larger GRCT on R2MR is indeed feasible.

In the current Scientific Report, we provide results from descriptive and efficacy analyses on the main study outcomes. Before we discuss these findings, we need to outline some important limitations. The

efficacy results from the pilot study must be interpreted with great caution, and with full consideration of these limitations. Some of these limitations are generic and apply to all pilot and feasibility studies in preparation for an RCT. We discuss these limitations first. Others are specific to the current pilot; we discuss these next.

First, it should be kept in mind that the existing scientific literature on pilot or feasibility studies in preparation for a full RCT or GRCT clearly recommends against using efficacy results from a pilot to determine whether a planned GRCT is feasible or should move forward. Pilot studies by definition have small samples and are not sufficiently-powered for hypothesis testing. Thus, "nonsignificant statistical tests—those that fail to achieve the largely arbitrary criterion of $p < .05$" [72] (p. 172) cannot be taken as "indicative of the poor feasibility of future planned research or as the need for "more research" before research can be scaled up."

Second, it should be kept in mind that the existing Scientific Literature on pilot or feasibility studies in preparation for a full RCT or GRCT also advises against using efficacy results from a pilot or feasibility study to determine in a preliminary fashion whether an intervention is beneficial or not (i.e., hypothesis testing) or how beneficial an intervention is (i.e., estimation of effect size). A pilot or feasibility study is not designed for these purposes. It usually has a small sample size and is simply not sufficiently powered for assessing intervention effects. The risk for Type II error (a false negative, in this case concluding that R2MR has no beneficial effect when in fact it does) is considerable in pilot studies. Furthermore, "it is possible, but *highly unlikely* [emphasis added], that the between group effect size (d) from a pilot study sample will provide a reasonable estimate of the population effect size (Δ), but that cannot be known based on the pilot data… This estimation problem has to do with the precision of d and its relation to sample size. Estimates become more precise with larger sample sizes" [73] (pp. 4–5). In the case of a significant intervention effect being detected in an underpowered pilot study, another caution should be taken in interpreting significant results. When sample size is small, the risk of Type I error (a false positive, concluding that R2MR has a beneficial effect when in fact it does not) is also inflated [74].

Third, pilot and feasibility studies have small, and likely unrepresentative, samples. Therefore, even if we were to ignore the problem of insufficient power to conduct hypothesis testing, we would still need to be very cautious in interpreting efficacy results given the limited generalizability of the pilot. Relatedly, if the pilot sample is not randomly selected (as in our case) and is possibly not representative of the larger population of interest (the NCM recruit population), significant results we see in the pilot may be unique to the pilot sample and may not be replicated in a larger GRCT.

As stated previously, in addition to the limitations of pilot and feasibility studies in general, there are specific limitations to the current pilot on R2MR that should add to the caution that must be exercised when interpreting efficacy results. As we briefly stated earlier in this report (and fully discuss in a separate report focusing on feasibility), due to delays in contracting, we could not use contractors blind to study allocation to administer the data collection sessions in the pilot. Furthermore, in order to be able to investigate problems with randomization, scheduling, and allocation during the pilot, we also did not attempt the blinding of the PI during the pilot. This introduces additional sources of bias into the pilot. To give one example, unblinding DRDC Toronto staff administering the data collection sessions could easily lead to greater effort being put into decreasing lost-to-follow-up (i.e., attrition) rate in the Intervention group, which can subsequently affect efficacy results.

Finally, as we stated previously in this Scientific Report, the platoons included in the pilot were those that differed significantly from all other platoons that complete their BMQ training at CFLRS. The

eight platoons included in the pilot paused their BMQ training and went home for three weeks during the Winter Break. Based on our prior research on the uptake of key learning objectives in R2MR in this population [19], [20], [26], [27] and as noted by others in military resilience RCTs [4], we suspect that the larger military training context can significantly influence/modify the presumed beneficial effects of interventions such as R2MR. Thus, including platoons that start their BMQ training then leave the BMQ training context for three weeks and finally return to finish their BMQ training further limits the generalizability of the pilot efficacy findings and raises the possibility that any "significant" findings that emerge in the pilot are unique to the pilot sample (i.e., risk of a false positive or Type I error).

Nevertheless, it is helpful to capture and publish descriptive and efficacy findings from pilot or feasibility studies. "Researchers have an ethical and scientific obligation to attempt publishing the results of every research endeavor" [75] (p. 6), as long as they are clear about limitations of their work and exercise caution in interpreting the results when such caution is warranted. In fact, a review of current practices in publishing pilot and feasibility studies [17] found that the vast majority (about 81%) included some form of hypothesis testing. Furthermore, without efficacy testing, it is not possible to rule out unanticipated harmful effects for interventions such as R2MR. Unanticipated harmful effects may seem unlikely but are not that rare in well-meaning medical, psychological or public health interventions [76], [77]. Finally, it should be noted that military resilience training programs like R2MR, such as the Comprehensive Soldier Fitness (CSF) program in the U.S. have at times drawn criticism from the larger research community for not fully outlining the empirical foundations of the interventions that have been put together and for implementing the interventions before well-designed randomized control trials have fully established the safety and efficacy of the interventions [78]. In the context of this ongoing debate as to how resilience interventions should be developed, tested, and implemented, it makes sense to capture all findings, include those from pilot and feasibility studies, in the public domain.

In the context of all the limitations we outlined above, we note here first and foremost that we did not see any evidence of harmful effects for R2MR in the analyses. Furthermore, the efficacy findings showed some interesting patterns. First, looking at the seven variables measuring attitudes and intentions towards mental health service use, we found a very consistent pattern of beneficial effects for R2MR by the second Follow-up. Additional analyses not shown here but available upon request from the first author show these between-group differences to be driven both by small movements within the Intervention group towards more favourable attitudes and intentions and by small movements towards less favourable attitudes and intentions within the Control group. These movements seem to start as we look at change over time from the Baseline to the first Follow-up, and continue as we look at change over time from Baseline to the Second Follow-up. These results are strikingly similar to those we obtained in a previous, small, controlled but non-randomized study on R2MR [25]. In that earlier study, we assigned two platoons each to one of three conditions: R2MR attitude change using a Video, R2MR attitude change using PowerPoint Slides (as in the current pilot), and a no R2MR Control condition. All platoons were assessed with the CAF-MHSUQ at Week 2 (prior to R2MR exposure for the two Intervention conditions) and four weeks later at Week 6 (after R2MR exposure for the two Intervention conditions), a similar timepoint to the first Follow-up in the current pilot. In that study, we also found consistent beneficial effects for the Slide R2MR condition over both the Video R2MR and the Control conditions. Similar to the current pilot findings, the between-group differences for the Slide R2MR versus the Control condition seemed to be driven both by small movements within the Slide condition towards more favourable attitudes and intentions and by small movements in the Control group towards less favourable attitudes and intentions, from Week 2 to Week 6 of the BMQ. While the consistency of the findings within the current pilot and the similarity of the current findings to previous findings are both reassuring and increase our confidence that the results are not due to chance, the definitive empirical test of the

beneficial effects of R2MR on MHSU attitudes and intentions must still await data from the larger GRCT on R2MR.

Second, looking at psychological health and psychological resilience outcomes, we find a consistent pattern of absence of beneficial effects at both Follow-up 1 and Follow-up 2. Given the Fidelity Checklist completed for the Intervention sessions showed perfect adherence to standard R2MR material, we do not suspect these findings to be driven by issues that relate to intervention fidelity. We did not previously examine the beneficial effects of R2MR on the outcomes included in this pilot; therefore, it is difficult to compare the current pilot findings to past R2MR research in this population. The absence of beneficial effects observed in this pilot are, however, consistent with the absence of beneficial effects on psychological health outcomes in the U.S. GRCT on mental health training during basic combat training [4]. Nevertheless, given the well-recognized limitations of pilot studies in hypothesis testing that we outlined above, and especially the very real chance for a false negative (Type II) error, the definitive test of the beneficial effects of R2MR on psychological health and resilience must also await data from the full GRCT.

There are many possible explanations for an absence of beneficial effects for psychological health and resilience outcomes. It is possible that the psychological distress experienced by most recruits during BMQ training is temporary and short-lived and may resolve on its own; an Intervention such as R2MR may not be needed and as a result, may not show beneficial effects when tested. Or, it may be that the key "active ingredients" of R2MR are not the right active ingredients. It may be that the Big 4 skills are not the right skills to use to improve psychological health and resilience outcomes in the BMQ context. Alternatively, the skills might be the right skills and they may be appropriate for the BMQ context; however, they may be inefficiently learned and practiced due to the physically and mentally taxing BMQ training the recruits are undergoing. When we look at the uptake of two key active ingredients in R2MR in this pilot study (mental health literacy and the use of the Big 4 skills), we find as expected, that the scores on both the mental health literacy and the use of the Big 4 skills are higher for the Intervention group compared to the Control group. However, these differences are not significant and are not large, suggesting perhaps that there may have been limited uptake of these critical learning objectives. Our previous research in which recruits were tested a day after exposure to R2MR and then several weeks later did show that there is limited uptake of the Big 4 skills in the BMQ context [19], [20], [26], [27]. Unfortunately, explanations for the absence of beneficial effects are speculative at best. Testing these speculative hypotheses is difficult, even in the fully-powered larger GRCT, let alone the current small pilot. A GRCT is designed and sufficiently powered to answer only two key questions: Is the intervention (R2MR) beneficial? And how large is the beneficial effect? GRCTs are not designed to answer why an intervention did not seem to have the presumed beneficial effects.

Third, looking at military performance outcomes, we found a mixed set of findings. For the Weapons Handling and Shooting tests, the Gas Hut, the FORCE test and the First Aid test, there were no beneficial effects detected. For the primary military performance outcome measure of BMQ graduation and the additional outcome of the 13-K March, we found a trend for beneficial effects. However, in both cases, the results were of great imprecision and uncertainty. The mixed set of results for R2MR's beneficial effects on military performance outcomes are similar to the mixed results reported in the one existing U.S. study on mental skills training during basic combat training [4]. The inconsistency of the pilot results and the large imprecision around the estimates for BMQ graduation and the 13-K March both argue for waiting until the larger GRCT to determine whether R2MR improves military performance outcomes, and if so, whether this beneficial effect applies to only some or all outcomes of interest.

3.6 Conclusion

In small pilot or feasibility studies, small sample size limits our ability for hypothesis testing. The common concern is increased Type II error in the case of obtaining non-significant results (i.e., a false negative, or erroneously concluding that R2MR has no beneficial effect when in fact it does). However, there is also a risk for increased Type I error (i.e., a false positive, or erroneously concluding the R2MR has beneficial effects even though it does not) in the case of obtaining significant results. The full empirical test of whether R2MR has beneficial effects must await data from the larger, sufficiently powered GRCT.

References

[1] Zamorski, M.A., Bennett, R.E., Boulos, D., Garber, B.G., Jetly, R., Sareen, J. The 2013 Canadian Forces mental health survey: background and methods. *Canadian Journal of Psychiatry* 2016; 61 (Supplement I): 10S–25S.

[2] Canadian Armed Forces. Road to Mental Readiness (R2MR). http://www.forces.gc.ca/en/caf-community-health-services-r2mr/index.page (Accessed July 14, 2017).

[3] Adler, A.B., Bliese, P.D., McGurk, D., Hoge, C.W., Castro, C.A. Battlemind debriefing and battlemind training as early interventions with soldiers returning from Iraq: Randomization by platoon. *Journal of Consulting and Clinical Psychology* 2009; 77(5): 928–40.

[4] Adler, A.B., Bliese, P.D., Pickering, M.A., Hammermeister, J., Williams, J., Harada, C., Csoka, L., Holiday, B., Ohlson, C. Mental skills training with basic combat training soldiers: A group-randomized trial. *Journal of Applied Psychology* 2015; 100(6): 1752–64.

[5] Adler, A.B., Williams, J., McGurk, D., Moss, A., Bliese, P.D. Resilience training with soldiers during basic combat training: randomisation by platoon. *Applied Psychology: Health and Wellbeing* 2015; 7(85–107).

[6] Castro, C., Adler, A.B., McGurk, D., Bliese, P.D. Mental health training with soldiers four months after returning from Iraq: Randomization by Platoon. *Journal of Traumatic Stress* 2012; 25: 376–83.

[7] Mulligan, K., Fear, N.T., Jones, N., Alvarez, H., Hull, L., Naumann, U., Wessely, S., Greenberg, N. Postdeployment Battlemind training for the U.K. armed forces: A cluster randomized controlled trial. *Journal of Consulting and Clinical Psychology* 2012; 80(3): 331–41.

[8] Cohn, A., Pakenham, K. Efficacy of a Cognitive-Behavioural program to improve psychological adjustment among soldiers in recruit training. *Military Medicine* 2008; 173(12): 1151–7.

[9] Williams, A., Hagerty, B.M., Andrei, A.-C., Yousha, S.M., Hirth, R.A., Hoyle, K.S. STARS: Stratgeies to assist Navy recruits' success. *Military Medicine* 2007; 172(9): 942–9.

[10] Williams, A., Hagerty, B.M., Yousha, S.M., Horrocks, J., Hoyle, K.S., Liu D. Psychosocial effects of the Boot Strap intervention in Navy recruits. *Military Medicine* 2004; 169(10): 814–20.

[11] Vanhove, A.J., Herian, M.N., Perez, A.L.U., Harms, P.D., Lester, P.B. Can resilience be developed at work? A meta-analytic review of resilience-building programme effectiveness. *Journal of Occupational and Organizational Psychology* 2015; 89(2): 278–307.

[12] Bliese, P.D., Adler, A.B., Castro, C.A. Research-based preventive mental health care strategies in the military. In: A.B. Adler, P.D. Bilese, & C.A. Castro, (Eds.). Deployment psychology: Evidence-based strategies to promote mental health in the military. Washington, DC: American Psychological Association; 2011: 103–24.

[13] Campbell, M.J., Walters, S.J. How to design, analyse, and report cluster randomised trials in medicine and health related research. West Sussex, U.K.: Wiley; 2014.

[14] Donner, A., Klar, N. Design and analysis of cluster randomization trials in health research. NY: New York, NY: Oxford University Press, 2000.

[15] Eldridge, S., Kerry, S. A practical guide to cluster randomized trials in health services research. West Sussex, U.K.: Wiley, 2012.

[16] Hayes, R.J., Moulton, L.H. Cluster randomized trials. Boca Raton, FL: Chapman and Hall/CRC; 2009.

[17] Arain, M., Campbell, M.J., Cooper, C.L., Lancaster, G.A. What is a pilot or feasibility study? A review of current practice and editorial policy. *BMC Medical Research Methodology* 2010; 10(67): 1–7.

[18] Blackler, K., Fikretoglu, D., Liu, A. Feasibility findings from a pilot Group Randomized Control Trial on the Road to Mental Readiness (R2MR) program, Defence Research and Development Canada, Scientific Report, DRDC-RDDC-2018-R003, 2018.

[19] Fikretoglu, D., Beatty, E., Liu, A. Optimizing R2MR at Basic Military Qualification (BMQ): Lessons learned from studies conducted between 2012 and 2014 and recommendations for implementation, Defence Research and Development Canada, Scientific Letter, DRDC-RDDC-2014-L244, 2014.

[20] Fikretoglu, D., Beatty, E., Liu, A. Comparing different versions of Road to Mental Readiness to determine optimal content: Testing instruction type, homework, and intelligence effects at two timepoints, Defence Research and Development Canada, Scientific Report, DRDC-RRDC-2014-R164, 2014.

[21] Fikretoglu, D., Lam, Q., Beatty, E. Recommendations to improve treatment fidelity and the uptake of R2MR material prior to a group randomized control trial, Defence Research and Development Canada, Letter Report, DRDC Toronto LR-2013-10055-1-1404HD0500, 2013.

[22] Jobidon, M.E., Blais, A.-R., Richards, K., Lam, Q., Fikretoglu, D. Attitudes des recrues Francophones de L'école de Leadership et de Recrues des Forces Canadiennes envers le recours aux services de santé mentale: Résultants et implications de la validation de la version française de questionnaire, Defence Research and Development Canada, Letter Report, DRDC Toronto LR-2013-10055-1-1404HD0500 to Directorate General Health Services, Directorate of Mental Health, 2013.

[23] Muller-Gass, A., Fikretoglu, D., Beatty, E., Vartanian, O., Paul, M. Assessing daytime sleepiness and fatigue in recruits at the Canadian Forces Leadership and Recruit School, Defence Research and Development Canada, Scientific Report, DRDC-RDDC-2015-R118, 2015.

[24] Lee, J., Fikretoglu, D., Blais, A.-R., Sudom, K., Beatty, E. Direct and mediated relationships of personality and individual difference characteristics with mental health service use intentions among Canadian military recruits. *Military Psychology* 2016; 28(6): 498–505.

[25] Fikretoglu, D., Liu, A., Blackler, K. Testing different methods to optimize change in mental health service use attitudes: Findings and recommendations for the Road to Mental Readiness (R2MR) program at basic Military Qualification, Defence Research and Development Canada, Scientific Report, DRDC-RDDC-2016-R025, 2016.

[26] Fikretoglu, D., Beatty, E., Liu, A., Guest, K., Bailey, S., (in press). Mental health education learning outcomes among Canadian Armed Forces recruits: Examining the effects of intelligence, instructor type, and instruction type. *Military Medicine*.

[27] Fikretoglu, D., Beatty, E., Liu, A. Comparing different versions of R2MR to determine optimal content: Testing instruction type, homework, and intelligence effects at two timepoints. Defence Research and Development Canada, Scientific Report, DRDC-RDDC-2014-R164, 2014.

[28] Fikretoglu, D., Lam, Q., Blais, A.-R. Attitudes towards mental health service use among CAF recruits during Basic Military Qualification: research findings and specific recommendations for improving the related R2MR content for recruits. Defence Research and Development Canada, Letter Report, DRDC Toronto LR-2013-10055-1-1404HD0500 to Directorate General Health Services, Directorate of Mental Health, 2013.

[29] Eldridge, S.M., Chan, C.L., Campbell, M.J., Bond, C.M., Hopewell, S., Thabane, L., Lancaster, G.A. CONSORT 2010 statement: extension to randomised pilot and feasibility trials. *BMJ Open* 2016; 355(i5239): 1–29.

[30] Spitzer, R.L., Kroenke, K., Williams, J.B. Group. PHQPCS. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. *Jama* 1999; 282(18): 1737–44.

[31] Kroenke, K., Spitzer, R.L. The PHQ-9: A new depression diagnostic and severity measure. *Psychiatric Annals* 2002; 32(9): 1–7.

[32] Kroenke, K., Spitzer, R.L., Williams, J.B. The PHQ-9. *Journal of General Internal Medicine* 2001; 16(9): 606–13.

[33] Cameron, I.M., Crawford, J.R., Lawton, K., Reid, I.C. Psychometric comparison of PHQ-9 and HADS for measuring depression severity in primary care. *British Journal of General Practice* 2008; 58(546): 32–6.

[34] Spitzer, R., Kroenke, K.J.W. Validation and utility of a self-report version of PRIME-MD: the PHQ Primary Care Study. *JAMA* 1999; 282: 1737–44.

[35] Kessler, R.C., Andrews, G., Colpe, L.J., Hiripi, E., Mroczek, D.K., Normand S.-L.T., Walters, E.E., Zaslavsky, A.M. Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychological Medicine* 2002; 32(06): 959–76.

[36] Kaplan, D.M., Smith, T., Coons, J.A validity study of the Subjective Unit of Discomfort (SUD) score. *Measurement and Evaluation in Counseling and Development* 1995; 27(4): 195–9.

[37] Blanc, S., Zamorski, M., Ivey, G., Edge, H.M., Hill, K. How much distress is too much on deployed operations? Validation of the Kessler Psychological Distress Scale (K10) for application in military operational settings. *Military Psychology* 2014; 26(2): 88–100.

[38] McFarlane, A., Hodson, S.E. Mental Health in the Australian Defence Force: 2010 ADF Mental Health Prevalence and Wellbeing Study: Report. Department of Defence, 2011.

[39] Slade, T., Johnston, A., Teesson, M., Whiteford, H.A., Burgess, P. The Mental Health of Australians 2. Report on the 2007 National Survey of Mental Health and Wellbeing. Department of Health and Ageing, Canberra. 2009.

[40] Tanner, B.A. Validity of global physical and emotional SUDS. *Applied psychophysiology and biofeedback* 2012; 37(1): 31–4.

[41] Spitzer, R.L., Kroenke, K., Williams, J.B., Löwe, B. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Archives of Internal Medicine* 2006; 166(10): 1092–7.

[42] Löwe, B., Decker, O., Müller, S., Brähler, E., Schellberg, D., Herzog, W., Herzberg, P.Y. Validation and standardization of the Generalized Anxiety Disorder Screener (GAD-7) in the general population. *Medical Care* 2008; 46(3): 266–74.

[43] Ruiz, M.A., Zamorano, E., García-Campayo, J., Pardo, A., Freire, O., Rejas, J. Validity of the GAD-7 scale as an outcome measure of disability in patients with generalized anxiety disorders in primary care. *Journal of affective disorders* 2011; 128(3): 277–86.

[44] Swinson, R.P. The GAD-7 scale was accurate for diagnosing generalised anxiety disorder. *Evidence Based Medicine* 2006; 11.

[45] Connor, K.M., Davidson, J.R. Development of a new resilience scale: The Connor-Davidson resilience scale (CD-RISC). *Depression and Anxiety* 2003; 18(2): 76–82.

[46] Campbell-Sills, L., Stein, M.B. Psychometric analysis and refinement of the connor–davidson resilience scale (CD-RISC): Validation of a 10-item measure of resilience. *Journal of Traumatic Stress* 2007; 20(6): 1019–28.

[47] Fikretoglu, D., Blais, A.-R., Lam, Q. Development and validation of a new Theory of Planned Behavior Questionnaire for mental health service use. Manuscript under revision, 2014.

[48] Ajzen, I. The theory of planned behavior. *Organizational Behavior and Human Decision Processes* 1991; 50: 179–211.

[49] Fikretoglu, D., McCreary, D.R., Blais, A.R., Sullivan-Kwantes, W., Jobidon, M.E. Developing a New Instrument to Assess Mental Healthcare Attitudes at the Beginning of the Military Career: Initial Psychometrics. Military and Veteran Health Research Forum; November 26–28, 2012; Kingston, Ontario, 2012.

[50] Fikretoglu, D., Sullivan-Kwantes, W., Richards, K. Identifying Mental Healthcare Attitudes at the Beginning of the Military Career: Implications for Current Mental Health Interventions. Military and Veteran Health Research Forum; November 26–28, 2012; Kingston, Ontario, 2012.

[51] Shipley, W.C., Gruber, C.P., Martin, T.A. Shipley-2 Manual. Los Angeles: Western Psychological Services, 2009.

[52] Crowne, D.P., Marlowe, D. A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology* 1960; 24(4): 349–54.

[53] Blais, A.R., Beatty, E., Fikretoglu, D., Blackler, K. Social desirability and selected mental health indices in Canadian Armed Forces recruits: Preliminary findings and recommendations to stakeholders. Defence Research and Development Canada, Scientific Letter, DRDC-RDDC-2015-L285, 2015.

[54] Thomas, P.R., Murphy, S.M., Hardy, L. Test of performance strategies: Development and preliminary validation of a comprehensive measure of athletes' psychological skills. *Journal of Sports Sciences* 1999; 17(9): 697–711.

[55] Woodcock, C. A multi-method examination of the processes and outcomes of IZOF interventions in competitive sport: implications for program design, delivery, and evaluation [Dissertation]: University of Birmingham, 2011.

[56] O'Connor, M., Casey, L., Clough, B. Measuring mental health literacy-a review of scale-based measures. *Journal of Mental Health* 2014; 23(4): 197–204.

[57] Wei, Y., McGrath, P.J., Hayden, J., Kutcher, S. Mental health literacy measures evaluating knowledge, attitudes and help-seeking: a scoping review. *BMC Psychiatry* 2015; 15: 291.

[58] Khan, S. Road to Mental Readiness – Evaluation report. SGR-2014-009, 2015.

[59] Liu, A., Fikretoglu, D. Power analysis for a proposed Group Randomized Control Trial (GRCT) on the Road to Mental Readiness (R2MR) program. Defence Research and Development Canada, Scientific Report, DRDC-RDDC-2014-R068, 2014.

[60] Liu, A., Fikretoglu, D., Blackler, K. the estimated sample size needed for the group randomized control trial on the Road to Mental Readiness: Updated results from a second power analysis. Defence Research and Development Canada, Scientific Letter, DRDC-RDDC-2017-L177, 2017.

[61] Efird, J. Blocked randomization with randomly selected block sizes. *International Journal of Environmental Research and Public health* 2011; 8: 15–20.

[62] Ivers, N.M., Halperin, I.J., Barnsley, J., Grimshaw, J.M., Shah, B.R., Tu, K., Upshur, R., Zwarenstein, M. Allocation techniques for balance at baseline in cluster randomized trials: a methodological review. *Trials* 2012; 120.

[63] Random Allocation Software Version 1.0 .http://mahmoodsaghaei.tripod.com/Softwares/randalloc.html. (Accessed April 15, 2016).

[64] Laird, N.M., Ware, J.H. Random-Effects Models for Longitudinal Data. *Biometrics (International Biometric Society* 1982; 38(4): 963–74.

[65] West, B.T., Welch, K.B., Galecki, A.T. Linear mixed models: A practical guide using statistical software. New York, NY: Chapman & Hall/CRC, 2007.

[66] Breslow, N.E, Clayton, D.G. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* 1993; 88(421): 9–25.

[67] Schulz, K.F., Grimes, D.A.L. Multiplicity in randomized trials I: endpoints and treatments. *Lancet* 2005; 365: 1591–5.

[68] Perneger, T.V. What's wrong with Bonferroni adjustments. *BMJ Open* 1998; 316: 11236–38.

[69] Streiner, D.L. Norman GRC – Correction for multiple testing—is there a resolution? *Chest* 2011; 140(1): 16–8.

[70] Cohen, J. Statistical power analysis for the behavioral sciences. Hillsdale, NJ: Lawrence Earlbaum Associates, 1988.

[71] SAS Institute Inc. Base SAS® 9.4 Procedures Guide. Cary, NC: SAS Institute Inc. 2015.

[72] Tickle-Degnen, L. Nuts and bolts of conducting feasibility studies. *American Journal of Occupational Therapy* 2013; 67: 171–6.

[73] Leon, A.C., Davis, L.L., Kraemer, H.C. The role and interpretation of pilot studies in clinical research. *Journal of Psychiatric Research* 2011; 45(5): 626–9.

[74] Colquhoun, D. An investigation of the false discovery rate and the misinterpretation of p-values. *Royal Society open science* 1.3 2014; 1: 140216.

[75] Thabane, L., Ma, J., Chu, R., Cheng, J., Ismaila, A., Rios, L.P., Robson, R., Thabane, M., Giangregorio L., Goldsmith, C.H. A tutorial on pilot studies: the what, why and how. *BMC Medical Research Methodology* 2010; 10(1): 1–10.

[76] Moyé, L.A. One-sided versus two-sided testing. In: L.A. Moyé, (Ed.). Statistical Reasoning in Medicine: The Intuitive P-Value Primer. 2nd edition. New York, NY: Springer; 2006.

[77] Bonell, C., Jamal, F., Melendez-Torres, G.J., Cummins, S. "Dark Logic": Theorising the harmful consequences of public health interventions. *Journal of Epidemiology and Community Health* 2015; 69: 95–8.

[78] Eidelson, R., Pilisuk, M., Soldz, S. The dark side of Comprehensive Soldier Fitness." 2011. https://www.psychologytoday.com/blog/dangerous-ideas/201103/the-dark-side-comprehensive-soldier-fitness. (Accessed July 14, 2017).

Annex A CONSORT Guidelines for Information to Include when Reporting a Pilot Trial

Table A.1: *CONSORT guidelines.*

Section/topic and item No	Standard checklist item	Extension for pilot trials
Title and abstract		
1a	Identification as a randomised trial in the title	Identification as a pilot or feasibility randomised trial in the title
1b	Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts)	Structured summary of pilot trial design, methods, results, and conclusions (for specific guidance see CONSORT abstract extension for pilot trials)
Introduction		
Background and objectives:		
2a	Scientific background and explanation of rationale	Scientific background and explanation of rationale for future definitive trial, and reasons for randomised pilot trial
2b	Specific objectives or hypotheses	Specific objectives or research questions for pilot trial
Methods		
Trial design:		
3a	Description of trial design (such as parallel, factorial) inclucing allocation ratic	Description of pilot trial design (such as parallel, factorial) including allocation ratio
3b	Important changes to methods after trial commencement (such as eligibility criteria), with reasons	Important changes to methods after pilot trial commencement (such as eligibility criteria), with reasons
Participants:		
4a	Eligibility criteria for participants	
4b	Settings and locations where the data were collected	
4c		How participants were identified and consented
Interventions:		
5	The interventions for each group with sufficient details to allow replication, including how and when they were actually administered	
Outcomes:		
6a	Completely defined prespecified primary and secondary outcome measures, including how and when they were assessed	Completely defined prespecified assessments or measurements to address each pilot trial objective specified in 2b, including how and when they were assessed
6b	Any changes to trial outcomes after the trial commenced, with reasons	Any changes to pilot trial assessments or measurements after the pilot trial commenced, with reasons
6c		If applicable, prespecified criteria used to judge whether, or how, to proceed with future definitive trial
Sample size:		
7a	How sample size was determined	Rationale for numbers in the pilot trial
7b	When applicable, explanation of any interim analyses and stopping guidelines	
Randomisation:		
Sequence generation:		
8a	Method used to generate the random allocation sequence	
8b	Type of randomisation; details of any restriction (such as blocking and block size)	Type of randomisation(s); details of any restriction (such as blocking and block size)
Allocation concealment mechanism:		
9	Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned	

Implementation:		
10	Who generated the random allocation sequence, enrolled participants, and assigned participants to interventions	
Blinding:		
11a	If done, who was blinded after assignment to interventions (eg, participants, care providers, those assessing outcomes) and how	
11b	If relevant, description of the similarity of interventions	
Analytical methods:		
12a	Statistical methods used to compare groups for primary and secondary outcomes	Methods used to address each pilot trial objective whether qualitative or quantitative
12b	Methods for additional analyses, such as subgroup analyses and adjusted analyses	Not applicable
Results		
Participant flow (a diagram is strongly recommended):		
13a	For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analysed for the primary outcome	For each group, the numbers of participants who were approached and/or assessed for eligibility, randomly assigned, received intended treatment, and were assessed for each objective
13b	For each group, losses and exclusions after randomisation, together with reasons	
Recruitment:		
14a	Dates defining the periods of recruitment and follow-up	
14b	Why the trial ended or was stopped	Why the pilot trial ended or was stopped
Baseline data:		
15	A table showing baseline demographic and clinical characteristics for each group	
Numbers analysed:		
16	For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups	For each objective, number of participants (denominator) included in each analysis. If relevant, these numbers should be by randomised group
Outcomes and estimation:		
17a	For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval)	For each objective, results including expressions of uncertainty (such as 95% confidence interval) for any estimates. If relevant, these results should be by randomised group
17b	For binary outcomes, presentation of both absolute and relative effect sizes is recommended	Not applicable
Ancillary analyses:		
18	Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing prespecified from exploratory	Results of any other analyses performed that could be used to inform the future definitive trial
Harms:		
19	All important harms or unintended effects in each group (for specific guidance see CONSORT for harms)	
19a		If relevant, other important unintended consequences

Annex B CONSORT Checklist

Table B.1: *CONSORT checklist.*

Item No	Standard checklist item	Extension for pilot trials	Page No
Title and abstract			
1a	Identification as a randomised trial in the title	Identification as a pilot or feasibility randomised trial in the title	Title page
1b	Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts)	Structured summary of pilot trial design, methods, results, and conclusions (for specific guidance see CONSORT abstract extension for pilot trials)	Page i
Introduction			
Background and objectives:			
2a	Scientific background and explanation of rationale	Scientific background and explanation of rationale for future definitive trial, and reasons for randomised pilot trial	Pages 1–2
2b	Specific objectives or hypotheses	Specific objectives or research questions for pilot trial	Pages 3–4
Methods			
Trial design:			
3a	Description of trial design (such as parallel, factorial) including allocation ratio	Description of pilot trial design (such as parallel, factorial) including allocation ratio	Pages 5–6
3b	Important changes to methods after trial commencement (such as eligibility criteria), with reasons	Important changes to methods after pilot trial commencement (such as eligibility criteria), with reasons	N/A
Participants:			
4a	Eligibility criteria for participants		Page 7
4b	Settings and locations where the data were collected		Page 7
4c		How participants were identified and consented	Page 7
Interventions:			
5	The interventions for each group with sufficient details to allow replication, including how and when they were		Page 8

	actually administered		
Outcomes:			
6a	Completely defined prespecified primary and secondary outcome measures, including how and when they were assessed	Completely defined prespecified assessments or measurements to address each pilot trial objective specified in 2b, including how and when they were assessed	Pages 8–13
6b	Any changes to trial outcomes after the trial commenced, with reasons	Any changes to pilot trial assessments or measurements after the pilot trial commenced, with reasons	N/A
6c		If applicable, prespecified criteria used to judge whether, or how, to proceed with future definitive trial	N/A
Sample Size:			
7a	How sample size was determined	Rationale for numbers in the pilot trial	Page 13
7b	When applicable, explanation of any interim analyses and stopping guidelines		
Randomization			
Sequence generation:			
8a	Method used to generate the random allocation sequence		Page 13
8b	Type of randomisation; details of any restriction (such as blocking and block size)	Type of randomisation(s); details of any restriction (such as blocking and block size)	Page 13
Allocation concealment mechanism:			
9	Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned		Pages 1–14
Implementation:			
10	Who generated the random allocation sequence, enrolled participants, and assigned participants to interventions		Pages 13–14
Blinding:			
11a	If done, who was blinded after assignment to interventions (e.g., participants, care providers, those		Page 14

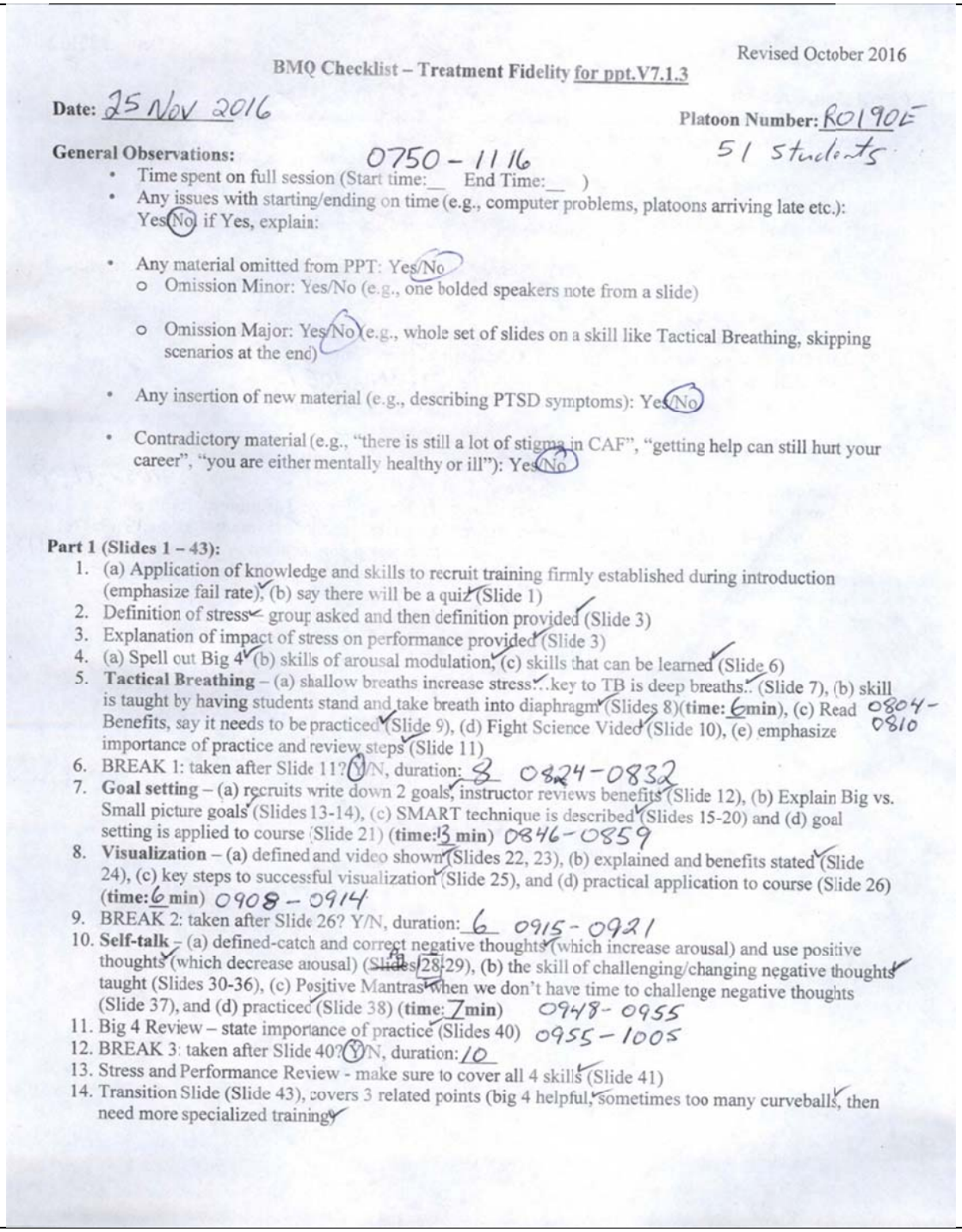
	assessing outcomes) and how		
11b	If relevant, description of the similarity of interventions		Page 8
Analytical methods:			
12a	Statistical methods used to compare groups for primary and secondary outcomes	Methods used to address each pilot trial objective whether qualitative or quantitative	Pages 14–15
12b	Methods for additional analyses, such as subgroup analyses and adjusted analyses	Not applicable	N/A
Results			
Participant flow (a diagram is strongly recommended):			
13a	For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analysed for the primary outcome	For each group, the numbers of participants who were approached and/or assessed for eligibility, randomly assigned, received intended treatment, and were assessed for each objective	Pages 16–17
13b	For each group, losses and exclusions after randomisation, together with reasons		Pages 16–17
Recruitment			
14a	Dates defining the periods of recruitment and follow-up		Page 18
14b	Why the trial ended or was stopped	Why the pilot trial ended or was stopped	Page 18
Baseline data:			
15	A table showing baseline demographic and clinical characteristics for each group		Page 19
Numbers analysed:			
16	For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups	For each objective, number of participants (denominator) included in each analysis. If relevant, these numbers should be by randomised group	Page 19, Table 2
Outcomes and estimation:			
17a	For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval)	For each objective, results including expressions of uncertainty (such as 95% confidence interval) for any estimates. If relevant, these	Pages 20–22

		results should be by randomised group	
17b	For binary outcomes, presentation of both absolute and relative effect sizes is recommended	Not applicable	Pages 22–23
Ancillary analyses:			
18	Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing prespecified from exploratory	Results of any other analyses performed that could be used to inform the future definitive trial	N/A
Harms:			
19	All important harms or unintended effects in each group (for specific guidance see CONSORT for harms)		N/A
19a		If relevant, other important unintended consequences	N/A
Discussion			
Limitations:			
20	Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses	Pilot trial limitations, addressing sources of potential bias and remaining uncertainty about feasibility	Pages 25–27
Generalisability:			
21	Generalisability (external validity, applicability) of the trial findings	Generalisability (applicability) of pilot trial methods and findings to future definitive trial and other studies	Pages 25–27
Interpretation:			
22	Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence	Interpretation consistent with pilot trial objectives and findings, balancing potential benefits and harms, and considering other relevant evidence	Page 27
22a		Implications for progression from pilot to future definitive trial, including any proposed amendments	N/A; Feasibility is discussed in a separate report
Other Information			
Registration:			
23	Registration number and name of trial	Registration number for pilot	Page 14

	registry	trial and name of trial registry	
Protocol:			
24	Where the full trial protocol can be accessed, if available	Where the pilot trial protocol can be ed, if available	Page 14
Funding			
25	Sources of funding and other support (such as supply of drugs), role of funders		Page 14
26		Ethical approval or approval by research review committee, confirmed with reference number	Page 14

Annex C Intervention Fidelity Checklists

Table C.1: Intervention fidelity checklists.



Revised October 2016

BMQ Checklist – Treatment Fidelity for ppt.V7.1.3

Date: 25 Nov 2016 **Platoon Number:** R0190E

51 students

General Observations: 0750 – 1116

- Time spent on full session (Start time:__ End Time:__)
- Any issues with starting/ending on time (e.g., computer problems, platoons arriving late etc.): Yes/No if Yes, explain:
- Any material omitted from PPT: Yes/No
  - Omission Minor: Yes/No (e.g., one bolded speakers note from a slide)
  - Omission Major: Yes/No (e.g., whole set of slides on a skill like Tactical Breathing, skipping scenarios at the end)
- Any insertion of new material (e.g., describing PTSD symptoms): Yes/No
- Contradictory material (e.g., "there is still a lot of stigma in CAF", "getting help can still hurt your career", "you are either mentally healthy or ill"): Yes/No

Part 1 (Slides 1 – 43):

1. (a) Application of knowledge and skills to recruit training firmly established during introduction (emphasize fail rate), (b) say there will be a quiz (Slide 1)
2. Definition of stress – group asked and then definition provided (Slide 3)
3. Explanation of impact of stress on performance provided (Slide 3)
4. (a) Spell out Big 4, (b) skills of arousal modulation, (c) skills that can be learned (Slide 6)
5. **Tactical Breathing** – (a) shallow breaths increase stress…key to TB is deep breaths. (Slide 7), (b) skill is taught by having students stand and take breath into diaphragm (Slides 8)(**time: 6 min**), (c) Read Benefits, say it needs to be practiced (Slide 9), (d) Fight Science Video (Slide 10), (e) emphasize importance of practice and review steps (Slide 11) 0804-0810
6. BREAK 1: taken after Slide 11? Y/N, duration: 8 0824-0832
7. **Goal setting** – (a) recruits write down 2 goals, instructor reviews benefits (Slide 12), (b) Explain Big vs. Small picture goals (Slides 13-14), (c) SMART technique is described (Slides 15-20) and (d) goal setting is applied to course (Slide 21) (**time: 3 min**) 0846-0859
8. **Visualization** – (a) defined and video shown (Slides 22, 23), (b) explained and benefits stated (Slide 24), (c) key steps to successful visualization (Slide 25), and (d) practical application to course (Slide 26) (**time: 6 min**) 0908 – 0914
9. BREAK 2: taken after Slide 26? Y/N, duration: 6 0915 - 0921
10. **Self-talk** – (a) defined-catch and correct negative thoughts (which increase arousal) and use positive thoughts (which decrease arousal) (Slides 28-29), (b) the skill of challenging/changing negative thoughts taught (Slides 30-36), (c) Positive Mantras when we don't have time to challenge negative thoughts (Slide 37), and (d) practiced (Slide 38) (**time: 7 min**) 0948 - 0955
11. Big 4 Review – state importance of practice (Slides 40) 0955 – 1005
12. BREAK 3: taken after Slide 40? Y/N, duration: 10
13. Stress and Performance Review - make sure to cover all 4 skills (Slide 41)
14. Transition Slide (Slide 43), covers 3 related points (big 4 helpful, sometimes too many curveballs, then need more specialized training)

Revised October 2016

Part II (Slides 44 – 57)

15. (a) Description of mental health continuum model – all colours/categories are described and the fact that there is movement in both directions is highlighted (Slide 44)
16. Behavioral Signs– using MHCM signs are described as they change along the continuum (Slide 46)
17. Healthy Coping: bounce back, peer support, buddy support (Slides 47-49)
18. When to seek help: (a) yellow to orange, (b) the six behavioral signs from the slide (Slide 50)
19. Seeking help (Slide 51): (a) 3 key points on slide are covered, (b) speaker notes starting with "Consider This" are covered
20. Barriers to timely help: 6 barriers and their respective challenges covered (Slides 52-54), together with speaker notes for each point
21. What goes on in treatment: all 4 points on (Slide 55) covered
22. Why seek treatment (Slide 56) all 3 key points on slide covered, together with speaker notes
23. That CAF treatment resources are Available, Accessible, Free (Slide 57)
24. BREAK 4: taken after Slide 57? Y/N, duration: 10 1041 - 1051

Part III (Slides 58 to end)

25. Course Recap, all points on (Slide 58) covered 1052-1102
26. (a) recruits divided into small groups, (b) scenarios and their questions are distributed, (c) 10 minutes given to groups (**time:10min**), (d) groups reconvene to go over scenarios (15 minutes) (**time:13min**) (e) for each scenario instructor ensures recruit answers capture key points included in speaker notes 1102 - 111
27. (a) Make sure to pose the question that forms the title of the slide and (b) cover Three Take home messages (Slide 67)
28. Tell them there is a quiz at some point soon

Table C.2: *Plattoon 189 Fidelity Checklist.*

Revised October 2016

BMQ Checklist – Treatment Fidelity for ppt.V7.1.3

Date: 24 Nov 2016 (Thu) **Platoon Number:** R0189 E

50 students

General Observations:

- Time spent on full session (Start time: 0854 End Time: 1207)
- Any issues with starting/ending on time (e.g., computer problems, platoons arriving late etc.): (Yes) No. if Yes, explain: Students 10 min late arriving at classroom
- Any material omitted from PPT: Yes (No)
  - Omission Minor: Yes (No) (e.g., one bolded speakers note from a slide)
  - Omission Major: Yes (No) (e.g., whole set of slides on a skill like Tactical Breathing, skipping scenarios at the end)
- Any insertion of new material (e.g., describing PTSD symptoms): Yes (No)
- Contradictory material (e.g., "there is still a lot of stigma in CAF", "getting help can still hurt your career", "you are either mentally healthy or ill"): Yes (No)

Part 1 (Slides 1 – 43):

1. (a) Application of knowledge and skills to recruit training firmly established during introduction (emphasize fail rate), (b) say there will be a quiz (Slide 1)
2. Definition of stress – group asked and then definition provided (Slide 3)
3. Explanation of impact of stress on performance provided (Slide 3)
4. (a) Spell out Big 4 (b) skills of arousal modulation, (c) skills that can be learned (Slide 6)
5. **Tactical Breathing** – (a) shallow breaths increase stress...key to TB is deep breaths.. (Slide 7), (b) skill is taught by having students stand and take breath into diaphragm (Slides 8)(**time:8 min**), (c) Read Benefits, say it needs to be practiced (Slide 9), (d) Fight Science Video (Slide 10), (e) emphasize importance of practice and review steps (Slide 11)
6. BREAK 1: taken after Slide 11? Y/N, duration: 9 0926-0935
7. **Goal setting** – (a) recruits write down 2 goals, instructor reviews benefits (Slide 12), (b) Explain Big vs. Small picture goals (Slides 13-14), (c) SMART technique is described (Slides 15-20) and (d) goal setting is applied to course (Slide 21) (**time:12 min**) 0948-1000
8. **Visualization** – (a) defined and video shown (Slides 22, 23), (b) explained and benefits stated (Slide 24), (c) key steps to successful visualization (Slide 25), and (d) practical application to course (Slide 26) (**time:6 min**) 1008-1014
9. BREAK 2: taken after Slide 26? Y/N, duration: 7 1015-1022
10. **Self-talk** – (a) defined-catch and correct negative thoughts (which increase arousal) and use positive thoughts (which decrease arcusal) (Slides 28-29), (b) the skill of challenging/changing negative thoughts taught (Slides 30-36), (c) Positive Mantras when we don't have time to challenge negative thoughts (Slide 37), and (d) practiced (Slide 38) (**time:7 min**) 1045-1052
11. Big 4 Review – state importance of practice (Slides 40)
12. BREAK 3: taken after Slide 40? Y/N, duration: 9 1053-1102
13. Stress and Performance Review - make sure to cover all 4 skills (Slide 41)
14. Transition Slide (Slide 43), covers 3 related points (big 4 helpful, sometimes too many curveballs, then need more specialized training)

Part II (Slides 44 – 57)

15. (a) Description of mental health continuum model – all colours/categories are described and the fact that there is movement in both directions is highlighted (Slide 44)
16. Behavioral Signs– using MHCM signs are described as they change along the continuum (Slide 46)
17. Healthy Coping: bounce back, peer support, buddy support (Slides 47-49)
18. When to seek help: (a) yellow to orange, (b) the six behavioral signs from the slide (Slide 50)
19. Seeking help (Slide 51): (a) 3 key points on slide are covered, (b) speaker notes starting with "Consider This" are covered
20. Barriers to timely help: 6 barriers and their respective challenges covered (Slides 52-54), together with speaker notes for each point
21. What goes on in treatment: all 4 points on (Slide 55) covered
22. Why seek treatment (Slide 56) all 3 key points on slide covered, together with speaker notes
23. That CAF treatment resources are Available, Accessible, Free (Slide 57)
24. BREAK 4: taken after Slide 57? (Y)N, duration: 7 1138-1145

Part III (Slides 58 to end)

25. Course Recap, all points on (Slide 58) covered 1145-1153
26. (a) recruits divided into small groups, (b) scenarios and their questions are distributed, (c) 10 minutes given to groups (**time:** 8 **min**), (d) groups reconvene to go over scenarios (15 minutes) (**time:** 12 **min**) (e) for each scenario instructor ensures recruit answers capture key points included in speaker notes 1153-1205
27. (a) Make sure to pose the question that forms the title of the slide and (b) cover Three Take home messages (Slide 67)
28. Tell them there is a quiz at some point soon

List of Symbols/Abbreviations/Acronyms/Initialisms

BMQ	Basic Military Qualification
CAF	Canadian Armed Forces
CAF-MHSUQ	Canadian Armed Forces Mental Health Service Use Questionnaire
CD-RISC	Connor-Davidson Resilience Scale
CFLRS	Canadian Forces Leadership and Recruit School
CMP	Chief of Military Personnel
CONSORT	Consolidated Standards of Reporting Trials
DGHS	Director General Health Services
DRDC	Defence Research and Development Canada
GAD-7	Generalized Anxiety Disorder Scale
GRCT	Group Randomized Control Trial
K-10	Kessler Psychological Distress Scale
MC-SDS	Marlowe-Crowne Social Desirability Scale
MHCM	Mental Health Continuum Model
MHL	Mental Health Literacy
MHSU	Mental Health Service Use
NCM	Non-Commissioned Member
OR	Odds Ratio
PHQ-9	Patient Health Questionnaire
PIN	Personal Identification Number
PLQ	Primary Leadership Qualification
RCT	Randomized Control Trial
RHQ	Recruit Health Questionnaire
R2MR	Road to Mental Readiness
SAS	Statistical Analysis Software
SUDS	Subjective Units of Distress Scale
TOPS	Test of Performance Strategies
U.K.	United Kingdom
U.S.	United States

DOCUMENT CONTROL DATA

(Security markings for the title, abstract and indexing annotation must be entered when the document is Classified or Designated)

1. ORIGINATOR (The name and address of the organization preparing the document. Organizations for whom the document was prepared, e.g., Centre sponsoring a contractor's report, or tasking agency, are entered in Section 8.)

DRDC – Toronto Research Centre
Defence Research and Development Canada
1133 Sheppard Avenue West
P.O. Box 2000
Toronto, Ontario M3M 3B9
Canada

2a. SECURITY MARKING (Overall security marking of the document including special supplemental markings if applicable.)

CAN UNCLASSIFIED

2b. CONTROLLED GOODS

NON-CONTROLLED GOODS
DMC A

3. TITLE (The complete document title as indicated on the title page. Its classification should be indicated by the appropriate abbreviation (S, C or U) in parentheses after the title.)

Testing the Efficacy of the Road to Mental Readiness (R2MR) Mental Health Education and Resilience Training Program During Basic Military Qualification: A Pilot Group Randomized Control Trial

4. AUTHORS (last name, followed by initials – ranks, titles, etc., not to be used)

Fikretoglu, D.; Liu, A.; Blackler, K.

5. DATE OF PUBLICATION (Month and year of publication of document.)

October 2018

6a. NO. OF PAGES (Total containing information, including Annexes, Appendices, etc.)

56

6b. NO. OF REFS (Total cited in document.)

78

7. DESCRIPTIVE NOTES (The category of the document, e.g., technical report, technical note or memorandum. If appropriate, enter the type of report, e.g., interim, progress, summary, annual or final. Give the inclusive dates when a specific reporting period is covered.)

Scientific Report

8. SPONSORING ACTIVITY (The name of the department project office or laboratory sponsoring the research and development – include address.)

DRDC – Toronto Research Centre
Defence Research and Development Canada
1133 Sheppard Avenue West
P.O. Box 2000
Toronto, Ontario M3M 3B9
Canada

9a. PROJECT OR GRANT NO. (If appropriate, the applicable research and development project or grant number under which the document was written. Please specify whether project or grant.)

Road to Mental Readiness

9b. CONTRACT NO. (If appropriate, the applicable number under which the document was written.)

10a. ORIGINATOR'S DOCUMENT NUMBER (The official document number by which the document is identified by the originating activity. This number must be unique to this document.)

DRDC-RDDC-2018-R158

10b. OTHER DOCUMENT NO(s). (Any other numbers which may be assigned this document either by the originator or by the sponsor.)

11a. FUTURE DISTRIBUTION (Any limitations on further dissemination of the document, other than those imposed by security classification.)

Public release

11b. FUTURE DISTRIBUTION OUTSIDE CANADA (Any limitations on further dissemination of the document, other than those imposed by security classification.)

12. ABSTRACT (A brief and factual summary of the document. It may also appear elsewhere in the body of the document itself. It is highly desirable that the abstract of classified documents be unclassified. Each paragraph of the abstract shall begin with an indication of the security classification of the information in the paragraph (unless the document itself is unclassified) represented as (S), (C), (R), or (U). It is not necessary to include here abstracts in both official languages unless the text is bilingual.)

Background: The Road to Mental Readiness (R2MR) program is the standard mental health education and resilience training program in the Canadian Armed Forces (CAF). The overall goal of R2MR training is to improve military performance, psychological health, resilience, and attitudes towards using mental health services. Since 2008, R2MR has been implemented throughout the military career and deployment cycles and thousands of military personnel have received R2MR. Like any other large-scale workplace mental health intervention, R2MR has to be tested for efficacy to see if it is achieving its program objectives. DRDC – Toronto Research Centre has been asked to conduct a Group Randomized Control Trial (GRCT) to test the efficacy of R2MR during military members' first exposure to the program, at Basic Military Qualification (BMQ).

Objective: A small pilot study was conducted in preparation for the larger GRCT, between October 31st, 2016 and February 8th, 2017. The primary objective of the pilot study was to assess the feasibility of the larger GRCT. The feasibility findings were summarized in an earlier report. The objectives of the current report are to provide descriptive and efficacy findings from the pilot study.

Methods: Eight Anglophone platoons were recruited for the study and randomized to an Intervention or a Delayed Intervention / Control condition. Three data collection sessions took place in Weeks 2, 5, and 9 of the BMQ (Baseline (T1), Follow-up 1 (T2) and Follow-up 2 (T3), respectively). At each data collection session, participants completed questionnaires assessing their psychological health and resilience as well as their attitudes and intentions towards mental health service use. Platoons randomized to the Intervention condition received R2MR in Week 2 of their BMQ (after the first data collection session); those randomized to the Control condition received R2MR in Week 9 of their BMQ (after the last data collection session). Performance outcomes were obtained for those participants who consented to data linkage to an administrative database at the Canadian Forces Leadership and Recruit School (CFLRS); data linkage was performed after the pilot study ended. Mixed effect models were obtained to examine efficacy. For continuous outcomes, we employed mixed linear models assuming random intercepts and slopes to account for platoon-level differences. For binary outcomes, we used generalized linear mixed models to assess individual-level differences while taking into account the platoon-level covariance.

Results: Out of a possible 427 Non-Commissioned Member (NCM) recruits, a total of 354 (82.90%) consented to participate in the study and completed T1 data collection. Of those original 354 participants, 296 completed T2 data collection (83.62%) and 278 completed T3 data collection (78.53%). A total of 267 participants (66.3%) provided consent to data linkage. There were no statistically significant differences between the two conditions on psychological health or resilience at Follow-up 1 or Follow-up 2. For some but not all of the performance outcomes, there was a trend toward beneficial effects. For attitudes and intentions towards mental health service use, there were consistent and statistically significant beneficial effects at Follow-up 2.

Conclusions: In this small pilot GRCT, we found mixed support for the presumed beneficial effects of R2MR. These results must be interpreted with great caution in the context of the well-recognized limitations of small pilot studies (e.g., the risk for Type I and II error) in general, and the specific limitations of this pilot study in particular.

Contexte: En route vers la préparation mentale (RVSM) est le programme de formation standard des Forces armées canadiennes (FAC) axé sur l'éducation en santé mentale et la résilience. L'objectif général du programme RVSM est d'améliorer le rendement, la santé mentale, la résilience et les attitudes à l'égard de l'utilisation des services de santé mentale par les militaires. Depuis 2008, le programme RVSM est mis en œuvre tout au long de la carrière militaire et des cycles de déploiement, et des milliers de militaires y ont déjà pris part. Comme dans toute autre intervention à grande échelle en santé mentale au travail, on a dû tester l'efficacité du programme RVSM afin de s'assurer qu'il atteignait bien ses objectifs. RDDC Toronto s'est vu confier le mandat de procéder à un essai contrôlé randomisé (ECR) par grappes pour évaluer l'efficacité du programme RVSM, lorsque les militaires y prennent part pour la première fois, durant la qualification militaire de base (QMB).

Objectif: Entre le 31 octobre 2016 et le 8 février 2017, on a réalisé une petite étude pilote afin de préparer la tenue de l'ECR par grappes à plus grande échelle. L'objectif premier de l'étude pilote était d'évaluer la faisabilité d'un ECR par grappes de plus grande envergure. Les conclusions quant à la faisabilité sont résumées dans un rapport précédent. Le présent rapport a pour objectif de fournir une description et les résultats quant à l'efficacité tirés de l'étude pilote.

Méthodes: Huit pelotons anglophones ont été recrutés pour l'étude et répartis aléatoirement soit dans le groupe expérimental, soit dans le groupe témoin (avec intervention différée). Trois séances de collecte de données ont eu lieu au cours des semaines 2, 5 et 9 de la QMB (référence [T1], suivi 1 [T2] et suivi 2 [T3], respectivement). Lors de chacune des séances de collecte de données, les participants ont rempli un questionnaire visant à évaluer leur santé mentale et leur résilience, ainsi que leurs attitudes et intentions à l'égard de l'utilisation des services de santé mentale. Les pelotons répartis aléatoirement du groupe expérimental ont pris part au programme RVSM durant la semaine 2 de leur QMB (après la première séance de collecte de données); ceux du groupe témoin ont pris part au programme RVSM durant la semaine 9 de leur QMB (après la dernière séance de collecte de données). On a obtenu des résultats de rendement chez les participants qui avaient consenti au couplage de ces données avec celles d'une base de données administrative de l'École de leadership et de recrues des Forces canadiennes (ELRFC). On a procédé au couplage des données une fois l'étude pilote terminée. Des modèles à effets mixtes ont été obtenus aux fins d'examen de l'efficacité. Dans le cas des résultats continus, nous avons eu recours à des modèles linéaires à effets mixtes avec ordonnée à l'origine et pentes aléatoires pour tenir compte des différences à l'échelle du peloton. Dans le cas des résultats binaires, nous avons utilisé des modèles linéaires généralisés mixtes pour évaluer les différences à l'échelle individuelle tout en tenant compte de la covariance à l'échelle du peloton.

Résultats: Sur 427 participants potentiels parmi les recrues militaires du rang (MR), 354 (82,90 %) ont accepté de prendre part à l'étude et terminé la collecte de données T1. Parmi les 354 participants du début, 296 ont terminé la collecte de données T2 (83,62 %) et 278 la collecte de données T3 (78,53 %). En tout, 267 participants (66,3 %) ont donné leur consentement au couplage des données. Au suivi 1 comme au suivi 2, on n'a noté aucune différence statistiquement significative entre les deux groupes sur le plan de la santé mentale ou de la résilience. Pour certains résultats de rendement, mais pas tous, une tendance vers des effets bénéfiques se dégage. Dans le cas des attitudes et des intentions à l'égard de l'utilisation des services de santé mentale, des effets bénéfiques ressortent de façon constante et statistiquement significative au suivi 2.

Conclusions: Au cours de la petite étude pilote prenant la forme d'un ECR par grappes, nous avons obtenu des résultats mitigés à l'appui des effets bénéfiques présumés du programme RVSM. Ces résultats doivent être interprétés avec la plus grande prudence compte tenu des

limites reconnues des petites études pilotes (p. ex. risque d'erreur de première ou de deuxième espèce) en général, et des limites propres à cette étude pilote en particulier.

13. KEYWORDS, DESCRIPTORS or IDENTIFIERS (Technically meaningful terms or short phrases that characterize a document and could be helpful in cataloguing the document. They should be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location may also be included. If possible keywords should be selected from a published thesaurus, e.g., Thesaurus of Engineering and Scientific Terms (TEST) and that thesaurus identified. If it is not possible to select indexing terms which are Unclassified, the classification of each should be indicated as with the title.)

Road to Mental Readiness; mental health education