Contemporary Clinical Trials Communications 19 (2020) 100631

Contents lists available at ScienceDirect

Contemporary Clinical Trials Communications

journal homepage: http://www.elsevier.com/locate/conctc

A data-driven examination of which patients follow trial protocol

Maren K. Olsen [a,b,*], Karen M. Stechuchak [a], Anna Hung [a,c], Eugene Z. Oddone [a,d], Laura J. Damschroder [f,g], David Edelman [a,d], Matthew L. Maciejewski [a,d,e]

[a] Durham Center of Innovation to Accelerate Discovery and Practice Transformation, Durham Veterans Affairs Health Care System, Durham, NC, USA
[b] Department of Biostatistics and Bioinformatics, Duke University, Durham, NC, USA
[c] DCRI, Duke University, Durham, NC, USA
[d] Division of General Internal Medicine, Department of Medicine, Duke University, Durham, NC, USA
[e] Department of Population Health Sciences, Duke University, Durham, NC, USA
[f] Ann Arbor VA HSR&D Center for Clinical Management Research, Ann Arbor, MI, USA
[g] VA PROVE QUERI, Ann Arbor, MI, USA

ARTICLE INFO

Keywords:
Veterans
Protocol adherence
Subgroup
Model-based recursive partitioning

ABSTRACT

Protocol adherence in behavioral intervention clinical trials is critical to trial success. There is increasing interest in understanding which patients are more likely to adhere to trial protocols. The objective of this study was to demonstrate the use of a data-driven approach to explore patient characteristics associated with the lowest and highest rates of adherence in three trials assessing interventions targeting behaviors related to lifestyle and risk for cardiovascular disease. Each trial included a common set of baseline variables. Model-based recursive partitioning (MoB) was applied in each trial to identify participant characteristics of subgroups characterized by these baseline variables with differences in protocol adherence. Bootstrap resampling was conducted to provide optimism-corrected c-statistics of the final solutions. In the three trials, rates of protocol adherence varied from 56.9% to 87.5%. Evaluation of heterogeneity of protocol adherence via MoB in each trial resulted in trees with 2–4 subgroups based on splits of 1–3 variables. In two of the three trials, the first split was based on pain in the past week, and those reporting lower pain were less likely to be adherent. In one of these trials, the second and third splits were based on education and employment, where those with lower education levels and who were employed were less likely to be adherent. In the third trial, the two splits were based on smoking status and then marriage status, where smokers who were married were least likely to be adherent. Optimism-corrected c-statistics ranged from 0.54 to 0.63. Model-based recursive partitioning can be a useful approach to explore heterogeneity in protocol adherence in behavioral intervention trials. An important next step would be to assess whether patterns hold in other similar studies and samples. Identifying subgroups who are less likely to be adherent to an intervention can help inform modifications to the intervention to help tailor the intervention to these subgroups and increase future uptake and impact.
Trial registration: ClinicalTrials.gov identifiers: NCT01828567, NCT02360293, and NCT01838226.

1. Introduction

Adherence to behavioral interventions that aim to prevent disease are challenging because patients are often asymptomatic and intervention goals are often focused on longer-term benefits (e.g., diabetes prevention) that may fail to motivate adherence in the short term [1]. Assessment of adherence in behavioral intervention trials can also be challenging because the interventions often have multiple components and it is not always clear, *a priori*, what degree of adherence participation is needed to demonstrate an intervention effect [2,3]. Despite these challenges, protocol adherence is increasingly reported as part of results in some behavioral intervention trials [4–8].

Assessment of protocol adherence is important because it is difficult to determine the causal effects of the intervention if adherence is suboptimal [2,9]. Null effects in a trial with poor protocol adherence could be due to provision of an ineffective intervention or provision of a behavioral intervention that would be effective if delivered and received as intended [3,9]. Assessing adherence is important to differentiating

* Corresponding author. Durham Center of Innovation to Accelerate Discovery and Practice Transformation, Durham Veterans Affairs Health Care System, Durham, NC, USA.
E-mail address: maren.olsen@duke.edu (M.K. Olsen).

https://doi.org/10.1016/j.conctc.2020.100631
Received 24 February 2020; Received in revised form 24 July 2020; Accepted 2 August 2020
Available online 13 August 2020

these two scenarios because the implications for intervention refinement are quite different.

Reporting of protocol adherence is also critical for improving the external validity of results [2,10] by revealing patient subgroups that have above- or below-average adherence to the treatment because such reporting can inform decisions on who to prioritize for future intervention. Deeper understanding of the barriers to intervention adherence may also inform how to modify or adapt the intervention to fit specific needs of patient subgroups.

Assessment of protocol adherence in behavioral-intervention prevention trials may involve consideration of both how the intervention was delivered and how the intervention was received. In 2004, the National Institutes of Health Behavioral Change Consortium (NIH BCC) created a framework for evaluating treatment fidelity, which refers to the "methodological strategies to monitor and enhance the reliability and validity of behavioral interventions" [9,11]. The NIH BCC framework evaluates treatment fidelity across five domains, including study design, provider training, treatment delivery (what providers teach), treatment receipt (what patients learn), and treatment enactment (what patients use) [9,11]. Due to the many components commonly involved in behavioral interventions and alternative modes (e.g., text, telephone, in-person individual, in-person group) by which interventions are delivered, defining a threshold for what is considered to be protocol-adherent can be difficult.

What remains understudied is how to use baseline participant characteristics to identify patient subgroups who are more likely to adhere to a behavioral intervention. The objective of this analysis was to illustrate the use of a data-driven approach to determine patient characteristics associated with lower and higher rates of adherence in three distinct behavioral-intervention prevention trials. These studies were jointly funded as a partnered research project with the Department of Veterans Affairs National Center for Health Promotion and Disease Prevention, yielding the possibility to establish a common set of baseline variables to identify subgroups. The three trials target lifestyle and self-management behavioral changes among different patient populations and levels of intervention intensity. This exploratory analysis illustrates the use of a data-driven method to understand heterogeneity in protocol adherence in three trials.

Methods

Trial design, participants, outcomes, and covariates

The three trials included in our analyses delivered coaching support via multiple modes: ACTIVATE (NCT01828567) and Stay Strong (NCT02360293) provided phone-based coaching, and Group Problem Solving (GPS; NCT01838226) provided coaching in-person and by phone. Each trial is described in detail elsewhere [12–14]. We briefly describe patient eligibility and each intervention below; further details are provided in eTable 1. Only participants randomized to the intervention arm of each trial are included in this analysis because adherence to the coaching component was the outcome of interest. One unique feature of these three trials is that the telephone coaches (n = 2) were the same for all three studies.

In the ACTIVATE trial, all randomized Veterans completed the Department of Veterans Affairs' (VA's) web-based health risk assessment (HRA) at baseline [12], which uses a proprietary risk modeling algorithm to provide patients with a "health age" based on lifestyle choices, family risk, biological values, and the degree to which lifestyle changes can lower their "health age." Veterans were eligible if they were enrolled in primary care at one of the three study sites and had at least one modifiable risk factor: body mass index (BMI) $\geq$30, current smoker, and/or less than 150 min of moderate/vigorous physical activity per week. Participants randomized to the intervention were scheduled to receive two telephone calls delivered by a health coach within one month after HRA completion. During the first call, ACTIVATE coaches reviewed HRA output with participants, explored participant preferences and values related to prevention, prioritized prevention topics, helped the participant choose a prevention program, collaboratively developed a SMART (smart, measurable, attainable, relevant, and timely) goal related to program enrollment, and assessed readiness and confidence to take the steps needed to enroll. One month after the initial coaching call, ACTIVATE coaches again called participants to review progress on his or her Prevention Action Plan. ACTIVATE coaches congratulated participants if they had enrolled in the prevention program or helped participants problem-solve any barriers to program enrollment if they had not yet enrolled.

The second trial, GPS, evaluated the effectiveness of a problem-solving therapy-based intervention for reducing cardiovascular risk among Veterans recruited from two VA primary care settings [14]. To be eligible, patients must have had no prior history of a cardiovascular event but were at elevated risk, as defined by having a Framingham Risk Score of at least 5%, with at least 2% of that risk reversible (e.g., smoking status can be reversed; age cannot). Patients randomized to the intervention were scheduled to attend six 90-min group sessions, conducted approximately monthly. The group sessions were designed such that the first three sessions focused on instructional content and the final three sessions focused on skill application. Between each group session, participants received an individualized telephone call (10–25 min each) from a trained health coach. The coaching calls focused on helping participants apply content from the preceding group session, develop personalized goals and strategies, and practice problem-solving skills to overcoming barriers to achieve their goals.

Finally, the goal of the Stay Strong randomized trial was to test the effect of adding telephone coaching to a mobile health intervention that included a physical activity monitor to improve and sustain levels of physical activity over 12 months among a national sample of US Veterans [13]. Veterans were eligible if they were an Operation Enduring Freedom (OEF)/Operation Iraqi Freedom (OIF)/Operation New Dawn (OND) Veteran under age 65, could identify a Veterans Health Administration (VHA) medical center and VHA health care provider responsible for his or her care, were interested in starting a physical activity program in the next 30 days, had access to a computer with an internet connection and a working Universal Serial Bus (USB) port, and had a smartphone running a compatible iOS or Android operating system. Participants randomized to the intervention arm received up to 3 calls with a coach within the first 9 weeks of the enrollment. During these calls, coaches assisted in developing goals and action plans, assisted in problem solving barriers to achieving these goals, and provided guidance on the Stay Strong app.

Outcomes

The outcome variable of adherence was defined uniquely for each trial, considering the nature of the intervention content, dose, and delivery. In ACTIVATE, intervention patients were defined as "adherent" if they completed both coaching calls. In GPS, participants were defined as "adherent" if they completed a combination of at least 8 of the 12 calls or in-person group sessions; the intervention developers felt that 8 of 12 contacts was the minimum required to assure an adequately potent dose of the intervention. Finally, in Stay Strong, intervention patients were defined as "adherent" if they completed at least two of the three planned coaching calls.

Baseline characteristics

Several participant demographic characteristics were commonly measured across all three trials and were included in our analyses: age, sex, race, highest level of education completed, marital status, financial distress, and employment status. Health-related measures included BMI and current smoking status [15]. A single item was used to characterize self-rated general health (excellent, very good, good, fair, poor) [16].

Several validated measures that were captured at baseline in all trials were also included as potential covariates for explaining heterogeneity in protocol adherence.

Additional measures that were standardized across all three trials include the Patient Activation Measure-13 (PAM-13), a 13-item measure that evaluates individuals' knowledge, skills, beliefs, and confidence for managing their health, and has demonstrated high construct validity [17,18]; the Patient Health Questionnaire-8 (PHQ-8), an eight-item measure that diagnoses and assesses severity of depression [19]; the Medical Outcomes Study (MOS) Sleep measure, a 6-item sleep measure that assesses important dimensions of sleep, such as initiation, maintenance, respiratory problems, quantity, perceived adequacy, and somnolence [20]; and the Alcohol Use Disorders Identification Test (AUDIT-C), a three-item measure of alcohol consumption [21]. Additionally, pain was measured from the single-item measure of pain from the 5-level version of EQ-5D (EQ-5D-5L), which is a widely used generic measure of health used to calculate quality-adjusted life years that measures five dimensions (mobility, self-care, usual activities, pain/-discomfort, and anxiety/depression) [22].

Statistical analyses

The intervention participants from each of the three trials were analyzed separately via a data-driven method known as model-based recursive partitioning (MoB). Given the differences in intervention content and mode of delivery, patient characteristics, and outcomes of interest, the trials were not pooled for analysis. Data-driven methods have been developed from statistical classification methods that lend themselves well to situations with many predictors with potentially complex interactions [23,24]. Model-based recursive partitioning [25] is one such method and was chosen because it can be used to discover prognostic factors associated with a dichotomous outcome. Other methods may only work for continuous outcomes or search for predictive factors in a heterogeneity of treatment effects context [23]. The basic premise of MoB is that, rather than one overall regression model, it may be possible to split the participants into subgroups based on the full set of available covariates, resulting in better fitting models for each respective subgroup, often comprising multiple covariates [26]. To assess whether or not a split on a covariate improves model fit, MoB tests for parameter instability across all values of the covariates; the value of the covariate associated with greatest parameter instability is used to define the first split (e.g. pain at a score of 1). A logistic regression model of the outcome variable of adherence is fit within each subgroup – (e.g., pain >1 versus pain ≤ 1). The process is repeated within each of the resulting subgroups until the best model fit is achieved, implicitly conducting variable selection. MoB yields a regression-based tree with each leaf or terminal node representing a subgroup experiencing differential adherence rates.

MoB was implemented via the mob function in the R package partykit version 1.2–8 [27] We specified a logistic regression model with the outcome of adherence and 15 potential baseline partitioning variables. Results are reported graphically as a regression-based tree.

Concordance statistics (c-statistics) were calculated to assess general model discrimination of the final solutions across the three trials. Additionally, for each trial, we conducted an internal validation by applying the MoB steps to 100 bootstrap samples, with c-statistics calculated for each sample. The resulting model from each bootstrap sample was also then applied to the original sample and the corresponding c-statistics calculated. The average difference between the two c-statistics provides an estimate of optimism (i.e., correcting for the original model performance being too optimistic) [28].

This study was approved by the Institutional Review Board (IRB) of the Durham VA Health Care System. The three trials were approved by the VHA Central IRB, as well as local IRB for Durham VA Medical Center, VA Western New York Healthcare System, and VA Ann Arbor Healthcare System.

Results

Participant characteristics

Across the three trials, participants were mostly male, had at least some college education or trade school, had average pain scores around 3–4, and reported at least fair general health (Table 1). Average participant age varied from 39.2 to 62.9 years of age, and the proportion employed varied from 32.2% to 66.3%.

Protocol adherence

Protocol adherence was 87.5% in the ACTIVATE trial, 71.3% in the Stay Strong trial and 56.9% in the GPS trial (Table 2), and was higher in the trials (ACTIVATE, Stay Strong) requiring fewer interactions to satisfy the definition of adherence. Evaluation of heterogeneity of protocol adherence via MoB in each trial resulted in a tree with 2–4 subgroups based on splits of 1–3 variables (Figs. 1–3). In two of the three trials (ACTIVATE, Stay Strong), the first split was based on pain in the past week.

In ACTIVATE, 92% of those with a pain score greater than 2 were adherent versus 77% of those with a pain score of 2 or less (Fig. 1). For this simple, two-subgroup solution, the c-statistic was 0.64 and the optimism-corrected c-statistic based on 100 bootstrap samples was 0.54.

The MoB solution for GPS resulted in 3 subgroups, defined by smoking status and marital status (Fig. 2). Participants who did not smoke had the highest rates of adherence, 69%. Among those who smoked, adherence rates were 52% for those who were not married and only 27% for those who were married. The c-statistic for this solution was 0.66 (optimism-corrected c-statistic = 0.60).

In Stay Strong, the algorithm found the best model fit with an initial split of the pain score at 1, with education and employment representing the next splits in the tree (Fig. 3), resulting in 4 subgroups. Among those with a pain score greater than 1, adherence rates were 90% for those who had at least a bachelor's degree, 81% for those without a bachelor's degree and unemployed, and 58% for those without a bachelor's degree and employed (Fig. 3). Fifty-five percent of those with a pain score of one or less were adherent to protocol. This solution had the highest c-statistic of 0.70 (optimism-corrected c-statistic = 0.63).

Discussion

In this study, we illustrated the use of a data-driven method, MoB, to identify subgroups of patients who might have had higher versus lower adherence to three behavioral prevention trials aimed at reducing cardiovascular risk and/or increasing physical activity. Though this method has been applied previously to assess heterogeneity of treatment effects, we apply it to examine variation in protocol adherence. While overall protocol adherence is increasingly reported in behavioral prevention trials, this analysis goes further to understand *heterogeneity* in protocol adherence [4–8].

Protocol adherence varied across the three trials. Unsurprisingly, protocol adherence was higher in the two trials requiring fewer interactions to be considered adherent. Participants reporting lowest levels of pain were least likely to be adherent in two of three trials. Possible explanations for this finding may be that these participants perceived less benefit from the interventions or that these participants were simply less available to answer the phone or attend sessions; conversely, patients with pain might have been strongly attracted to interventions that had heavy telephone components thereby requiring less travel to attend sessions. Although none of the three interventions addressed pain specifically, patients with higher pain scores may have had higher adherence due to secondary gain achieved from these mostly telephone-based coaching interventions. Patients with pain may realize health benefits from participation because all three interventions addressed behavioral aspects of risk reduction that may also improve

Table 1
Baseline characteristics of intervention participants, by trial.

	ACTIVATE N = 208	GPS N = 202	Stay Strong N = 178
Age, mean (SD)[a]	55.3 (12.7)	62.9 (11.1)	39.2 (8.4)
Male, No. (%)	172 (82.7)	181 (89.6)	131 (73.6)
Non-Hispanic white race, No. (%)	95 (45.7)	127 (62.9)	117 (65.7)
Highest level of education, No. (%)			
High school or less	39 (18.8)	32 (15.8)	16 (9.0)
Some college, Associate's degree, or trade school	112 (53.8)	125 (61.9)	90 (50.6)
Bachelor's degree or higher	57 (27.4)	45 (22.3)	72 (40.4)
Married/living as married, No. (%)	96 (46.2)	109 (54.0)	120 (67.4)
Employed full or part time, No. (%)[a]	79 (38.2)	65 (32.2)	118 (66.3)
Financial status, No. (%)[a]			
After paying the bills, you still have enough money for special things that you want.	70 (33.7)	87 (43.3)	81 (45.5)
You have enough money to pay the bills, but little spare money to buy extra or special things.	81 (38.9)	85 (42.3)	75 (42.1)
You have money to pay the bills, but only because you have cut back on things.	37 (17.8)	17 (8.5)	12 (6.7)
You are having difficulty paying the bills, no matter what you do.	20 (9.6)	12 (6.0)	10 (5.6)
PAM score, mean (SD)	62.4 (12.7)	59.3 (11.3)	70.8 (16.2)
PHQ score, mean (SD) [a]	6.8 (5.4)	5.1 (4.8)	8.0 (6.0)
MOS Score, mean (SD)	61.2 (21.3)	66.4 (23.0)	60.9 (19.7)
Pain past week, mean (SD)[b]	4.4 (2.7)	3.9 (2.8)	3.3 (2.3)
AUDIT-C, mean (SD)[a]	2.4 (2.8)	2.7 (2.6)	2.6 (2.1)
Body mass index, mean (SD)[a]	33.7 (6.6)	30.2 (5.3)	31.3 (6.2)
General health, No. (%)			
Excellent	13 (6.3)	26 (12.9)	8 (4.5)
Very Good	43 (20.7)	69 (34.2)	36 (20.2)
Good	85 (40.9)	67 (33.2)	71 (39.9)
Fair	50 (24.0)	35 (17.3)	50 (28.1)
Poor	17 (8.2)	5 (2.5)	13 (7.3)
Current smoker of cigarettes or other tobacco, No. (%)	88 (42.3)	88 (43.6)	50 (28.1)

Note: SD = standard deviation, GPS = Group Problem Solving, PAM = Patient Activation Measure, PHQ = Patient Health Questionnaire, MOS = Medical Outcomes Study, AUDIT-C = Alcohol Use Disorders Identification Test – Consumption. Group percentages may not add to 100% due to rounding.

[a] Missing data: age (n = 1, Stay Strong), employment (No. = 1, ACTIVATE), financial status (No. = 1, GPS), PHQ score (No. = 4, Stay Strong), AUDIT-C (No. = 2, GPS), body mass index (No. = 7, GPS). Percentage calculations exclude observations with missing data from the denominator.

[b] Pain is measured on a 0–10 scale, with 0 representing no pain.

Table 2
Adherence Details from each Trial.

Study	Definition of Adherence	No. Adherent/No. In study sample (%)
ACTIVATE	Completed both coaching calls	182/208 (87.5)
GPS	Completed at least 8 (of 12) coaching calls or group sessions	115/202 (56.9)
Stay Strong	Completed at least 2 (of 3) coaching calls	127/178 (71.3)

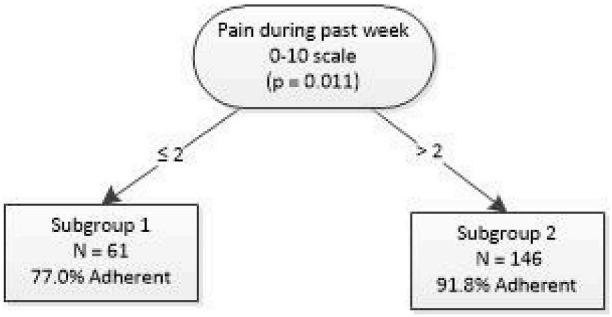

Fig. 1. ACTIVATE trial: MoB final solution. We used the default value of statistical significance for the fluctuation tests (alpha = 0.05). Instead of specifying a Bonferroni correction (which would have altered the statistical significance to 0.05/15), we chose to post prune by Akaike's Information Criteria fit index and set the minimum node sample size as 30. Finally, we specified maxLM-type test as the fluctuation test for ordered factor variables. All other control parameters were kept at their default values. The final sample size was 207 due to missing data on candidate baseline characteristics.

pain (e.g., setting SMART goals, increasing physical activity). If this finding is replicated across other behavioral intervention studies, including a brief pain assessment as part of baseline could help target extra resources for intervention fidelity or tailoring for inclusion/exclusion criteria.

More robust statistical analyses, like MoB, beyond traditional intention to treat-based analyses can provide deeper understanding of intervention non-adherence, which may help point to ways to adapt interventions to specific subgroups. For example, in the GPS trial, smokers who were married were least likely to be adherent to the intervention and represent a subgroup that VA could focus on to better understand their barriers and how to modify the intervention. Furthermore, data-driven methods, such as MoB, have an advantage of being able to characterize multivariable subgroups and are well suited to situations with many predictors with potentially complex interactions and little *a priori* knowledge concerning which subgroups may exhibit common patterns [23]. The methods also do not rely on *a priori* cut-points or categorizations of continuous predictors. Yet, with that flexibility comes the need for validation of the data-driven findings. Other analyses using data-driven methods to identify predictors of low intervention adherence have emphasized the importance of validation [29]. To reflect the potential for overfitting in our analyses, we provided optimism-corrected c-statistics derived via bootstrap resampling. The final optimism-corrected c-statistics were all less than 0.7, indicating weak discriminative ability. Given the complicated nature of adherence to behavioral interventions, it is not surprising that discriminative ability was weak [2,3]. Nonetheless, an important next step would be to assess the performance of these methods and explore whether similar patterns hold in other studies and samples.

We had a unique opportunity to compare results across three different intervention trials because baseline measures were common across all three trials. However, a limitation of our analyses is that the interventions were delivered to different Veteran populations and differed in intensity, dose, and mode of delivery, which required different definitions of adherence across the trials. Yet, the interventions were all focused on improving modifiable risk factors to decrease cardiovascular risk, such as increasing physical activity, evaluated similar baseline variables, and the telephone coaches were the same for all three trials. Each study's definition of adherence was agreed upon by each of the respective study teams based on theories related to their particular intervention. A further limitation is that because all trials and interventions were focused on the Veteran population, this study may not generalize beyond Veterans.

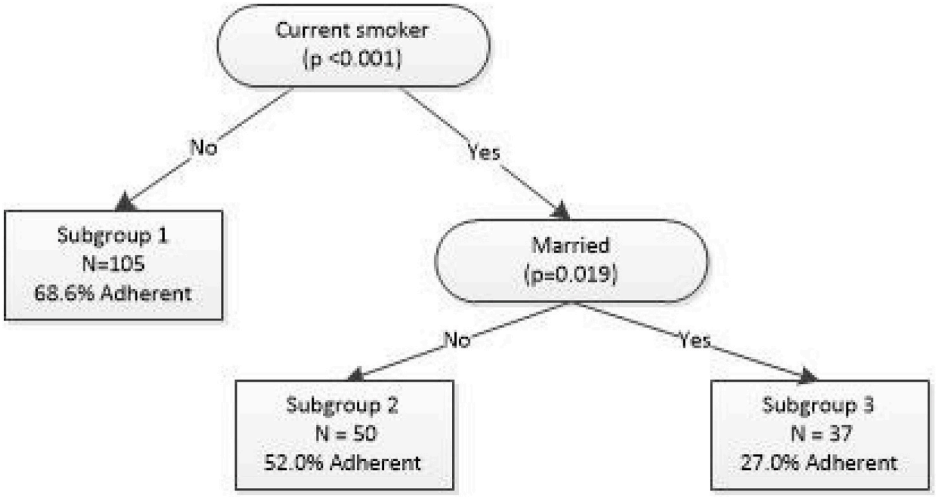

Fig. 2. GPS trial: MoB final solution. We used the default value of statistical significance for the fluctuation tests (alpha = 0.05). Instead of specifying a Bonferroni correction (which would have altered the statistical significance to 0.05/15), we chose to post prune by Akaike's Information Criteria fit index and set the minimum node sample size as 30. Finally, we specified maxLM-type test as the fluctuation test for ordered factor variables. All other control parameters were kept at their default values. The final sample size was 192 due to missing data on candidate baseline characteristics.

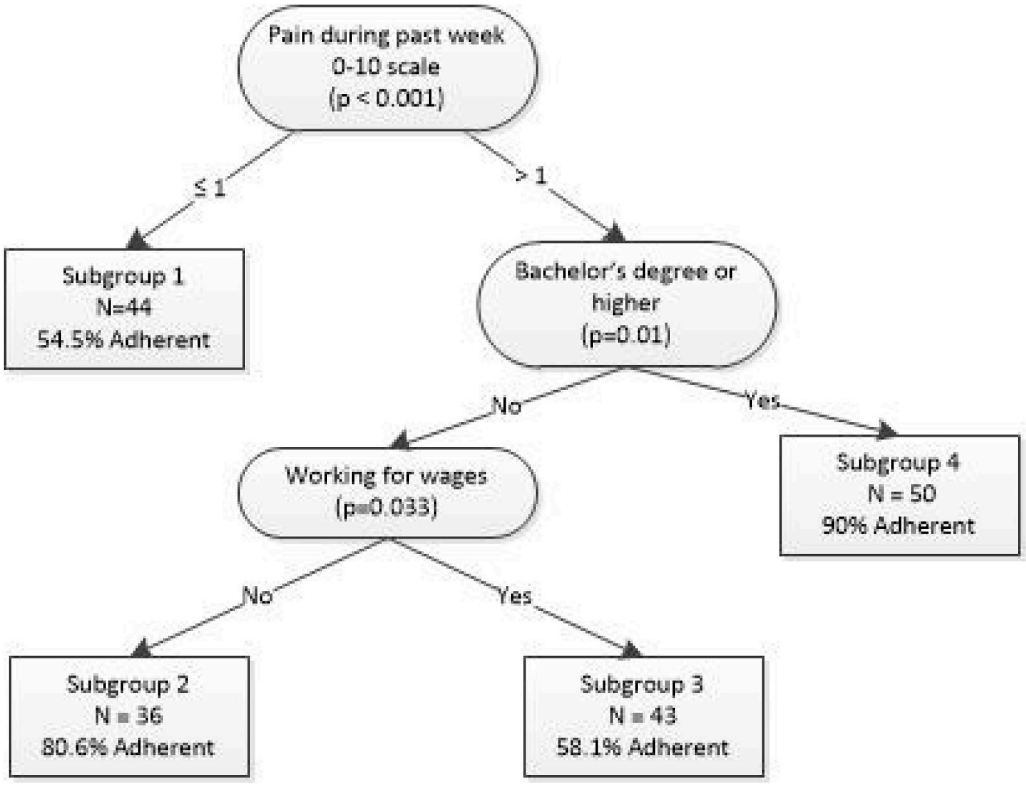

Fig. 3. Stay Strong trial: MoB final solution. We used the default value of statistical significance for the fluctuation tests (alpha = 0.05). Instead of specifying a Bonferroni correction (which would have altered the statistical significance to 0.05/15), we chose to post prune by Akaike's Information Criteria fit index and set the minimum node sample size as 30. Finally, we specified maxLM-type test as the fluctuation test for ordered factor variables. All other control parameters were kept at their default values. The final sample size was 173 due to missing data on candidate baseline characteristics.

While each behavioral intervention trial included details on how intervention fidelity would be measured during the trial (e.g., audio recording), it is also important to understand the types of patients who are more likely to and less likely to adhere to the behavioral intervention. In the future, analyses of heterogeneity of patient adherence to protocol, using data-driven approaches like MoB, should be considered to help assess impact of adherence, including dimensions described by the NIH BCC framework. Methods presented in this paper illustrate the strength of data-driven methods to help inform real-world implementation and adaptations as the VA scales up its programs using coaching sessions to improve cardiovascular risk factors in Veterans.

Declaration of competing interest

All authors concur with this submission. This manuscript has not been submitted to another journal nor published elsewhere. Apart from Dr. Maciejewski, no authors have financial or commercial conflicts of interest. Dr. Maciejewski owns Amgen stock due to his spouse's employment.

Acknowledgements

This work was supported by HSR&D funding (CRE 12–306, RCS 10–391) and by the Durham Center of Innovation to Accelerate Discovery and Practice Transformation (ADAPT), (CIN 13–410) at the Durham VA Health Care System. The views represented in this article

represent those of the authors and not those of the VA or the United States Government.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.conctc.2020.100631.

References

[1] M. Segar, No Sweat: How the Simple Science of Motivation Can Bring You a Lifetime of Fitness: 506, AMACOM div american mgmt assn, 2015.

[2] D. McGee, F. Lorencatto, K. Matvienko-Sikar, E. Toomey, Surveying knowledge, practice and attitudes towards intervention fidelity within trials of complex healthcare interventions, Trials 19 (1) (2018) 504–506, https://doi.org/10.1186/s13063-018-2838-6 [doi].

[3] H. Walton, A. Spector, I. Tombor, S. Michie, Measures of fidelity of delivery of, and engagement with, complex, face-to-face health behaviour change interventions: a systematic review of measure quality, Br. J. Health Psychol. 22 (4) (2017) 872–903, https://doi.org/10.1111/bjhp.12260 [doi].

[4] K.P. McNamara, S.L. O'Reilly, J. George, et al., Intervention fidelity for a complex behaviour change intervention in community pharmacy addressing cardiovascular disease risk, Health Educ. Res. 30 (6) (2015) 897–909, https://doi.org/10.1093/her/cyv050 [doi].

[5] E. Toomey, J. Matthews, D.A. Hurley, Using mixed methods to assess fidelity of delivery and its influencing factors in a complex self-management intervention for people with osteoarthritis and low back pain, e015452-015452, BMJ Open 7 (8) (2017), https://doi.org/10.1136/bmjopen-2016-015452 [doi].

[6] T. Swindle, J.P. Selig, J.M. Rutledge, L. Whiteside-Mansell, G. Curran, Fidelity monitoring in complex interventions: a case study of the WISE intervention, 53-2. eCollection 2018, Arch. Publ. Health 76 (2018), https://doi.org/10.1186/s13690-018-0292-2 ([doi]).

[7] Y. Fukuoka, W. Haskell, F. Lin, E. Vittinghoff, Short- and long-term effects of a mobile phone app in conjunction with brief in-person counseling on physical activity among physically inactive women: the mPED randomized clinical trial, JAMA Netw Open 2 (5) (2019), e194281, https://doi.org/10.1001/jamanetworkopen.2019.4281.

[8] S. Balducci, V. D'Errico, J. Haxhi, et al., Effect of a behavioral intervention strategy on sustained change in physical activity and sedentary behavior in patients with type 2 diabetes: the IDES_2 randomized clinical trial, J. Am. Med. Assoc. 321 (9) (2019) 880–890, https://doi.org/10.1001/jama.2019.0922 [doi].

[9] B. Borrelli, The assessment, monitoring, and enhancement of treatment fidelity in public health clinical trials, J. Publ. Health Dent. 71 (s1) (2011) S52–S63, https://doi.org/10.1111/j.1752-7325.2011.00233.x ([doi]).

[10] O. O'Shea, R. McCormick, J.M. Bradley, B. O'Neill, Fidelity review: a scoping review of the methods used to evaluate treatment fidelity in behavioural change interventions, Phys. Ther. Rev. 21 (3–6) (2016) 207–214, https://doi.org/10.1080/10833196.2016.1261237, doi: 10.1080/10833196.2016.1261237.

[11] A.J. Bellg, B. Borrelli, B. Resnick, et al., Enhancing treatment fidelity in health behavior change studies: best practices and recommendations from the NIH behavior change consortium, Health Psychol. 23 (5) (2004) 443–451, https://doi.org/10.1037/0278-6133.23.5.443 ([doi]).

[12] E.Z. Oddone, L.J. Damschroder, J. Gierisch, et al., A coaching by telephone intervention for veterans and care team engagement (ACTIVATE): a study protocol for a hybrid type I effectiveness-implementation randomized controlled trial, Contemp. Clin. Trials 55 (2017) 1–9, doi: S1551-7144(16)30350-0 [pii].

[13] L.R. Buis, F.A. McCant, J.M. Gierisch, et al., Understanding the effect of adding automated and human coaching to a mobile health physical activity app for Afghanistan and Iraq veterans: protocol for a randomized controlled trial of the stay strong intervention, JMIR Res Protoc 8 (1) (2019), e12526, https://doi.org/10.2196/12526 ([doi]).

[14] J.A. Nieuwsma, L.O. Wray, C.I. Voils, et al., A problem-solving intervention for cardiovascular disease risk reduction in veterans: protocol for a randomized controlled trial, Contemp. Clin. Trials 60 (2017) 42–50, doi: S1551-7144(16)30179-3 [pii].

[15] D.E. Nelson, D. Holtzman, J. Bolen, C.A. Stanwyck, K.A. Mack, Reliability and validity of measures from the behavioral risk factor surveillance system (BRFSS), Sozial- Präventivmed. 46 (Suppl 1) (2001) 3.

[16] J. Ware Jr., M. Kosinski, S.D. Keller, A 12-item short-form health survey: construction of scales and preliminary tests of reliability and validity, Med. Care 34 (3) (1996) 220–233, https://doi.org/10.1097/00005650-199603000-00003 [doi].

[17] J.H. Hibbard, E.R. Mahoney, J. Stockard, M. Tusler, Development and testing of a short form of the patient activation measure, Health Serv. Res. 40 (6 Pt 1) (2005) 1918–1930, doi: HESR438 [pii].

[18] J.H. Hibbard, E.R. Mahoney, R. Stock, M. Tusler, Do increases in patient activation result in improved self-management behaviors? Health Serv. Res. 42 (4) (2007) 1443–1463, doi: HESR669 [pii].

[19] K. Kroenke, T.W. Strine, R.L. Spitzer, J.B. Williams, J.T. Berry, A.H. Mokdad, The PHQ-8 as a measure of current depression in the general population, J. Affect. Disord. 114 (1–3) (2009) 163–173, https://doi.org/10.1016/j.jad.2008.06.026 ([doi]).

[20] R.D. Hays, S.A. Martin, A.M. Sesti, K.L. Spritzer, Psychometric properties of the medical outcomes study sleep measure, Sleep Med. 6 (1) (2005) 41–44, doi: S1389-9457(04)00129-7 [pii].

[21] K. Bush, D.R. Kivlahan, M.B. McDonell, S.D. Fihn, K.A. Bradley, The AUDIT alcohol consumption questions (AUDIT-C): an effective brief screening test for problem drinking. ambulatory care quality improvement project (ACQUIP). alcohol use disorders identification test, Arch. Intern. Med. 158 (16) (1998) 1789–1795, https://doi.org/10.1001/archinte.158.16.1789 [doi].

[22] A.S. Pickard, M.C. De Leon, T. Kohlmann, D. Cella, S. Rosenbloom, Psychometric comparison of the standard EQ-5D to a 5 level version in cancer patients, Med. Care 45 (3) (2007) 259–263, https://doi.org/10.1097/01.mlr.0000254515.63841.81 [doi].

[23] I. Lipkovich, A. Dmitrienko, B.R. D'Agostino S, Tutorial in biostatistics: data-driven subgroup identification and analysis in clinical trials, Stat. Med. 36 (1) (2017) 136–196, https://doi.org/10.1002/sim.7064 [doi].

[24] C. Strobl, J. Malley, G. Tutz, An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests, Psychol. Methods 14 (4) (2009) 323–348, https://doi.org/10.1037/a0016973 [doi].

[25] H. Seibold, A. Zeileis, T. Hothorn, Model-based recursive partitioning for subgroup analyses, Int. J. Biostat. 12 (1) (2016) 45–63, https://doi.org/10.1515/ijb-2015-0032 [doi].

[26] A. Zeileis, T. Hothorn, K. Hornik, Model-based recursive partitioning, J. Comput. Graph Stat. 17 (2) (2008) 492–514, https://doi.org/10.1198/106186008X319331, doi: 10.1198/106186008X319331.

[27] T. Hothorn, A. zeileis, partykit: a modular toolkit for recursive partytioning in R, J. Mach. Learn. Res. 16 (2015) 3905–3909.

[28] F.E. Harrell Jr., K.L. Lee, D.B. Mark, Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors, Stat. Med. 15 (4) (1996) 361–387, doi: 10.1002/(SICI)1097-0258(19960229)15:43.0.CO;2-4 [pii].

[29] J. Wallert, E. Gustafson, C. Held, et al., Predicting adherence to internet-delivered psychotherapy for symptoms of depression and anxiety after myocardial infarction: machine learning insights from the U-CARE heart randomized controlled trial, J. Med. Internet Res. 20 (10) (2018), e10754, https://doi.org/10.2196/10754 [doi].