SMARTp: A SMART design for non-surgical treatments of chronic periodontitis with spatially-referenced and non-randomly missing skewed outcomes

Journal:	*Biometrical Journal*
Manuscript ID	bimj.201900027.R2
Wiley - Manuscript type:	Research Paper
Date Submitted by the Author:	n/a
Complete List of Authors:	Xu, Jing; Duke-NUS Medical School, Centre for Quantitative Medicine ; Children's Medical Research Institute, Cancer Data Science Bandyopadhyay, Dipankar; Virginia Commonwealth University, Department of Biostatistics Mirzaei, Sedigheh; St. Jude Children's Research Hospital, Department of Biostatistics Michalowicz, Bryan; HealthPartners Institute for Education and Research Chakraborty, Bibhas; Duke-NUS Graduate Medical School, National University of Singapore, Centre for Quantitative Medicine; National University of Singapore, Department of Statistics and AppliedProbability
Keywords:	Dynamic treatment regimes, inverse probability weighting, method of moments, periodontitis, skew-normal, skew-t, SMART

Biometrical Journal **60** (2019) 10, zzz–zzz / DOI: 10.1002/bimj.201900000

SMARTp: A SMART design for non-surgical treatments of chronic periodontitis with spatially-referenced and non-randomly missing skewed outcomes

Jing Xu [1,2], **Dipankar Bandyopadhyay** [3], **Sedigheh Mirzaei Salehabadi** [4], **Bryan Michalowicz** [5], and **Bibhas Chakraborty**[*1,6,7]

[1] Centre for Quantitative Medicine and Programme in Health Services and Systems Research, Duke-NUS Medical School, Singapore

[2] Cancer Data Science, Children's Medical Research Institute, The University of Sydney, Westmead, NSW, Australia

[3] Department of Biostatistics, Virginia Commonwealth University, Richmond, VA, USA

[4] Department of Biostatistics, St. Jude Children's Research Hospital, Memphis, TN, USA

[5] HealthPartners Institute, Bloomington, MN, USA

[6] Department of Statistics and Applied Probability, National University of Singapore, Singapore

[7] Department of Biostatistics and Bioinformatics, Duke University, Durham, NC, USA

Received zzz, revised zzz, accepted zzz

This paper proposes dynamic treatment regimes (DTRs) as effective individualized treatment strategies for managing chronic periodontitis. The proposed DTRs are studied via `SMARTp` – a two-stage sequential multiple assignment randomized trial (SMART) design. For this design, we propose a statistical analysis plan and a novel cluster-level sample size calculation method that factors in typical features of periodontal responses, such as non-Gaussianity, spatial clustering, and non-random missingness. Here, each patient is viewed as a cluster, and a tooth within a patient's mouth is viewed as an individual unit inside the cluster, with the tooth-level covariance structure described by a conditionally autoregressive structure. To accommodate possible skewness and tail behavior, the tooth-level clinical attachment level (CAL) response is assumed to be skew-t, with the non-randomly missing structure captured via a shared parameter model corresponding to the missingness indicator. The proposed method considers mean comparison for the regimes with or without sharing an initial treatment, where the expected values and corresponding variances or covariance for the sample means of a pair of DTRs are derived by the inverse probability weighting and method of moments. Simulation studies are conducted to investigate the finite-sample performance of the proposed sample size formulas under a variety of outcome-generating scenarios. An `R` package `SMARTp` implementing our sample size formula is available at the Comprehensive R Archive Network for free download.

Key words: Dynamic treatment regimes; inverse probability weighting; method of moments; periodontitis; skew-normal; skew-t; SMART

1 Introduction

Chronic periodontitis (CP) is a serious form of periodontal disease (PD), and if left untreated, continues to remain a major cause of adult tooth loss (Eke *et al.*, 2012). It is a highly prevalent condition, affecting almost half of the US adults above the age of 30 years (Thornton-Evans *et al.*, 2013). Periodontal disease (PD) may exhibit significant comorbidity, including diabetes, cardiovascular complications, respiratory illnesses, etc. (Grossi *et al.*, 1997; Wang *et al.*, 2009; Beck *et al.*, 2001). Dental hygienists consider the

*Corresponding author: Centre for Quantitative Medicine and Programme in Health Services and Systems Research, Duke-NUS Medical School, Academia, 20 College Road, 06-41, Singapore, 169856. E-mail: `bibhas.chakraborty@duke-nus.edu.sg`

Clinical **A**ttachment **L**evel (CAL), rounded to the nearest millimeter, as the most important biomarker to measure the severity of PD (Nicholls, 2003). The CAL refers to the amount of lost periodontal ligament fibers, with the severity categorized as Slight/Mild: 1-2mm, Moderate: 3-4mm, and Severe: $\geq$ 5mm, according to the American Academy of Periodontology (AAP) 1999 guidelines (Wiebe and Putnins, 2000).

The treatment options for periodontitis range from oral hygiene, to scaling and root planing (SRP), to SRP with adjunctive treatments, and eventually to surgeries, when the severity increases. Recommended bi-annual (basic) dental cleaning, polishing, and professional flossing which we enjoy sitting in a comfortable reclining chair in a dental clinic only removes plaque (bacterial colony) and tartar above the gum line, and are not considered as effective procedures for treating gum diseases. Often, early stages of CP are effectively treated via non-surgical means. Benefits of non-surgical treatment of CP includes shorter recovery time due to less-invasive techniques, reduced discomfort than surgery, and fewer dietary restrictions leading to improved quality of life, post procedure. When disease has progressed significantly, *deep* cleaning, a two-step process consisting of scaling and (tooth) root planing (SRP), is often recommended as the gold standard (Herrera, 2016) for thorough plaque removal, at or below the gum line. Yet, bacteria may still exist under the gum line following an SRP. Hence, the dentist or oral hygienist may also recommend various supplementary procedures, or adjuncts (such as locally-delivered antimicrobials, systemic antimicrobials, non-surgical lasers, etc) following SRP to treat chronically deep gum pockets. Also, patients with periodontal co-morbidity who responds sub-optimally to SRP may benefit from adjunctive therapy (Porteous and Rowe, 2014). However, current evidence suggests that the use of these adjuncts as stand-alone treatments does not lead to any clinical benefits of treating CP, compared to SRP alone (Azarpazhooh *et al.*, 2010; Sgolastra *et al.*, 2012). Hence, they are usually recommended in conjunction to SRP.

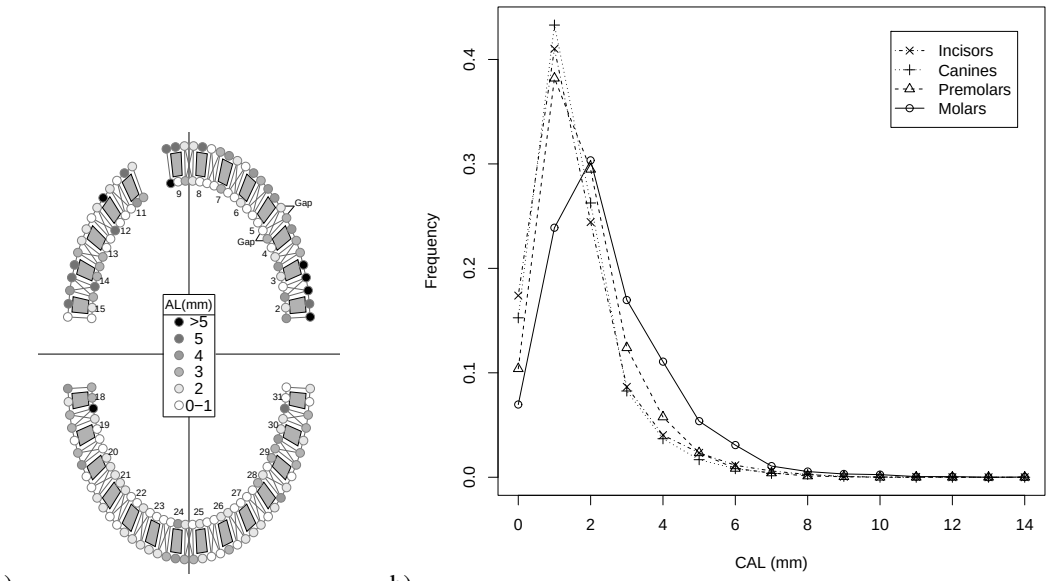
In 2011, the Council on Scientific Affairs of the American Dental Association (ADA) resolved to develop a Clinical Practice Guideline (CPG) for the non-surgical treatment (Smiley *et al.*, 2015) of CP with SRP, with or without adjuncts, based on a systematic literature review. The panel found 0.5 mm average improvement in CAL with SRP, while the combination of SRP with assorted adjuncts resulted in (average) CAL improvement of 0.2-0.6 mm, over SRP alone. However, comparison among the adjuncts were not conducted. Recently, a systematic network meta-analyses (NMA) of the adjuncts in 74 studies from the CPG revealed none of them to be statistically (significantly) superior to the other (John *et al.*, 2017). However, the NMA ranked SRP + doxycycline hyclate gel (a local antimicrobial) as the best non-surgical treatment of CP, compared to SRP alone. Throughout the years, lasers have revolutionized oral care, with reported advantages in minimizing tissue damages, swelling, and bleeding, leading to high patient acceptance. However, its clinical efficacy, both as an alternative, or adjuvant to SRP (Liu *et al.*, 1999; Zhao *et al.*, 2014) remains inconclusive, with inconsistent evidence derived from underpowered clinical studies (Porteous and Rowe, 2014); see also the April 2011 statement (Workgroup, 2011) by the AAP.

Although randomized clinical trials (RCTs) and subsequent meta-analyses continue to remain the *de facto* in understanding effectiveness of a treatment (or intervention) over others, conducting a successful RCT comes with its own bag of limitations, which includes issues with patient recruitment and retention, escalating costs, educating the wider public, etc. This has led to the development of more *patient-centric* approaches, primarily through adaptive interventions, or dynamic treatment regimes (DTRs) (Murphy *et al.*, 2001; Murphy, 2003; Robins, 2004) under the umbrella of precision medicine (Garcia *et al.*, 2013), where the focus changed to decision-making for the individual, or subgroups, rather than average-response based traditional RCT treatment comparisons. DTRs involve multistage sequential interventions, according to the patients' evolving characteristics (such as an individual patient's response, or adherence) at each subsequent treatment stage. The treatment types are repeatedly adjusted over time to match an individual's need in order to achieve optimal treatment effect. They are very appealing in managing chronic diseases that require long-term care (Lavori and Dawson, 2004; Murphy and McKay, 2004). Examples of DTRs from various clinical areas include alcoholism (Breslin *et al.*, 1998), smoking (Chakraborty *et al.*, 2010),

drug abuse (Brooner and Kidorf, 2002), depression (Untzer *et al.*, 2001), hypertension (Glasgow *et al.*, 1989), etc. However, in the field of oral health and CP, a DTR proposal to mitigate the aforementioned issues related to RCTs seem to be non-existent. For example, in the treatment of CP, one may develop a simple adaptive strategy of continuing the SRP (in the later stages) among the SRP responders-only group (in the first stage), while subjecting the non-responders to 'SRP + adjuncts' at later stages.

A special class of designs, called **S**equential **M**ultiple **A**ssignment **R**andomized **T**rial, or SMART designs (Murphy, 2005; Lei *et al.*, 2012), are popularly used to study DTRs (Chakraborty and Moodie, 2013; Ertefaie *et al.*, 2015). SMART designs involve randomizing patients to available treatment options at the initial stage, followed by re-randomizing at each subsequent stage of some or all of the patients to treatments available at that stage. These re-randomizations and set of treatment options depend on how well the patient responded to the previous treatments (Ghosh *et al.*, 2016). Contrary to the standard RCT that takes a 'one size fits all' single intervention approach, the SMART advances the RCT by following the same individual over sequential randomizations (i.e., ordered interventions) with the underlying series and order of interventions depicting a real-life setting, that can drastically affect eventual outcomes. Although Murphy (2005) proposed the general SMART design framework, the treatment was restricted to individual-level outcomes. However, in our motivating clinical discipline of treating CP (and also in some behavioral intervention research), interventions are delivered at the group or cluster (individual) level, while the CAL responses are available at the cluster sub-unit level (teeth) level. Although sample size formulas and SMART design implementations under clustered outcomes setting are available (Ghosh *et al.*, 2016; NeCamp *et al.*, 2017), they only focus on regimes that do not share an initial treatment. Furthermore, they do not account for other data complications typical to PD studies, such as presence of (i) non-Gaussian (skewed and thick-tailed), (ii) non-randomly missing, and (iii) spatially-referenced CAL responses (Reich *et al.*, 2013). For example, consider the motivating **G**ullah **A**frican **A**merican **D**iabetic, or GAAD study, which recorded the extent of PD in a Type-2 diabetic Gullah-speaking African American population from the coastal South Carolina sea-islands (Fernandes *et al.*, 2009). For illustration, panel (a) in Figure 1 describes the measurement locations and sample data for a random subject, while panel (b) plots the density histogram of the CAL for the four tooth-types from the GAAD dataset, revealing considerable right-skewness. Furthermore, PD being the major cause of adult tooth-loss, it is likely that patients with higher level of CAL (and CP) exhibit a higher proportion of missing teeth, and hence this missingness mechanism is non-ignorable (Reich *et al.*, 2013). Also, CP and PD are hypothesized to be spatially-referenced, i.e., proximally located teeth usually have similar disease status than distally located ones. Ignoring the features (i)–(iii) in constructing any SMART design for CP may lead to imprecise estimates of the desired parameters. It is important to note here that the Ghosh *et al.* (2016) approach of a clustered SMART design considers traditional clustering (sub-units within a cluster) of Gaussianly distributed continuous responses, and excludes spatial clustering and other features.

In this paper, we set forward to address the aforementioned limitations in developing a list of plausible DTRs for treating CP. We cast this into a two-stage SMART design framework for CP outcomes that exhibit (i)–(iii), and present an analysis plan and sample size calculations for (a) detecting a postulated effect size of a single treatment regime, and (b) detecting a postulated difference between two treatment regimes with or without a shared initial treatment. The tooth-level covariance structure describing spatial association is modeled by a conditionally autoregressive process (Reich and Bandyopadhyay, 2010). To accommodate possible skewness and tail behavior, the tooth-level CAL responses are assumed to have skew-t (Azzalini and Capitanio, 2003a) errors, with the non-randomly missing CAL values imputed via a shared parameter model corresponding to the missingness indicator. The proposed method considers mean comparison for the regimes with or without sharing an initial treatment, where the expected values and corresponding variances or covariance of the effect size of the treatment regimes are derived by the inverse probability weighting (IPW) techniques (Robins *et al.*, 1994a), and method of moments. Note that the proposed methodology is not restricted to CP trials only. It can also be applied (or extended) to other



a) b)

Figure 1 CAL data. Panel (a) shows the observed CAL for a patient with a missing incisor, where the shaded boxes represent teeth, the circles represent sites, and gray lines represent neighbour pairs that connects adjacent sites on the same tooth and sites that share a gap between teeth. "Gap" in the figure indicates, for example, the four sites in the gap between teeth # 4 and 5. The tooth numbers are indicated, and excludes the 4 third-molars: 1, 16, 17 32. The vertical and horizontal lines separate the mouth into four quadrants, with the molars (# 2-3, 14-15, 18-19, 30-31), premolars (# 4-5, 12-13, 20-21, 28-29), canines (# 6, 12, 22, 27) and incisors (# 7-10, 23-26). Panel (b) presents the frequency density plot of the CAL (rounded to the nearest mm) for each tooth type from the GAAD dataset.

SMARTs involving DTRs for treating other health conditions (e.g. infectious diseases) for which outcomes may be skewed and spatially correlated, with some outcome data possibly being non-randomly missing.

The rest of the paper is organized as follows. Section 2 introduces eight potential treatments, and the corresponding DTRs that constitute the 2-stage SMART design for CP. Section 3 presents the theoretical framework and a sample size calculation method under this SMART design, incorporating the aforementioned features typical to PD data. Section 4 investigates the finite-sample performance of the proposed sample size calculation method using synthetic data generated under various settings. Section 5 demonstrates the implementation of the `R` function `SampleSize.SMARTp` for calculating sample sizes, also available at the `GitHub` link `https://github.com/bandyopd/SMARTp`. Finally, the paper ends with a discussion in Section 6. Supplementary materials, consisting of detailed derivations are relegated to the Appendix.

2 A Proposed SMART design involving DTRs for treating CP

In this section, we propose dynamic treatment regimes (DTRs) for treating CP, which are studied via a SMART design. A list of possible treatments consist of the treatment initiation steps: (1) Oral hygiene

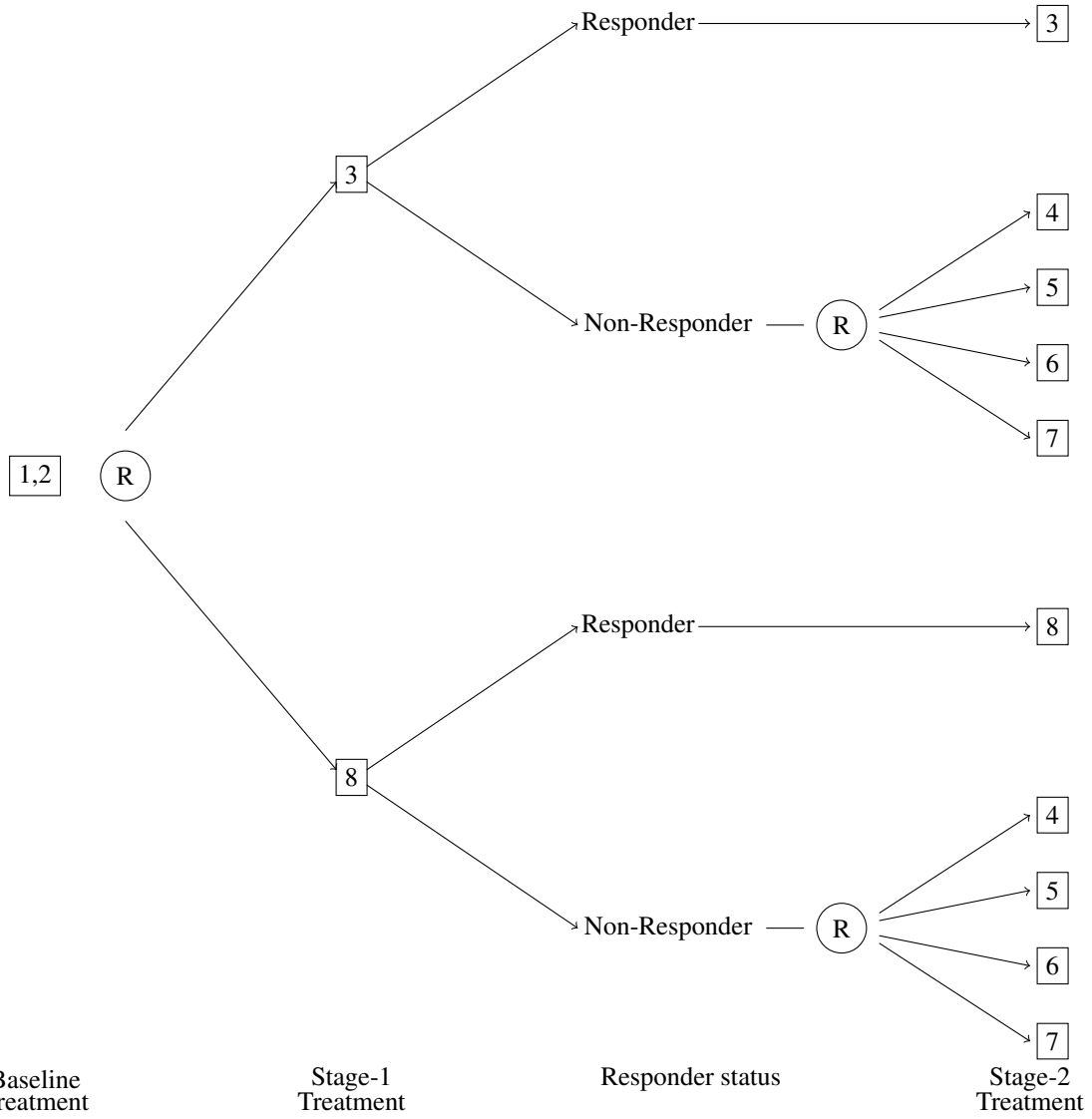


Figure 2 A SMART design schematic diagram for developing DTRs for treating chronic periodontitis. R = randomization, 1 = oral hygiene instruction, 2 = education on risk reduction, 3 = scaling and root planing (SRP), 4 = SRP with local antimicrobial therapy, 5 = SRP with systemic antimicrobial therapy, 6 = SRP with photodynamic therapy, 7 = SRP with systemic sub-antimicrobial-dose doxycycline (SDD), 8=laser therapy.

instruction, and (2) Education on risk reduction. This is followed by (3) SRP, or more advanced non-surgical treatments that combine SRP with adjunctive therapy, as summarized in the systematic review of Smiley *et al.* (2015), such as (4) SRP with local antimicrobial therapy, (5) SRP with systemic antimicrobial therapy, (6) SRP with photodynamic therapy, which uses lasers, but only to activate an antimicrobial agent, (7) SRP with systemic subantimicrobial-dose doxycycline (SDD), and finally, (8) Laser. The corresponding SMART design for developing DTRs is presented in Figure 2. Note that the number of potential DTRs are not limited to Figure 2. More details are presented in Section 6.

Oral hygiene is primarily used for prevention and initial therapy, especially during early stage of periodontitis. At the beginning of the proposed trial, each participant has to attend the treatment initiation steps (1) and (2) before any randomization. Note that while SRP is the accepted gold-standard, the role of laser therapy, though advantageous in targeting the diseased area precisely and accurately, still remains controversial as a standard of care. In this paper, we develop our SMART design, with a primary focus on comparing the DTRs starting with either SRP (# 3) or Laser therapy (# 8). At the initial stage, each participant is randomly allocated to either treatment 3 or 8. We propose a DTR that matches a patient's need in achieving similar outcome as SRP with adjuncts, though at a lower cost. Each possible treatment regime can have more than one path of treatment according to each patient's evolving response. The patients who respond to the initial treatment continue the same treatment at the 2nd stage of the trial. The patients who do not respond to treatment 3 are randomly allocated to one of the treatments 4–7 in the 2nd stage. Similarly, for patients allocated to the laser arm (treatment 8), the non-responders will also have the provision of being randomly allocated to one of treatments 4–7 in the 2nd stage. The randomization probabilities calculated at both the initial and the final stages of our SMART design are presented in Section 3.3. The primary final outcome measure is the recorded and rounded tooth-level CAL. The possible paths are listed below.

- Path 1: '1', '2', '3', '3'
- Path 2: '1', '2', '3', '4'
- Path 3: '1', '2', '3', '5'
- Path 4: '1', '2', '3', '6'
- Path 5: '1', '2', '3', '7'
- Path 6: '1', '2', '8', '8'
- Path 7: '1', '2', '8', '4'
- Path 8: '1', '2', '8', '5'
- Path 9: '1', '2', '8', '6'
- Path 10: '1', '2', '8', '7'

This leads to eight different DTRs (d_1–d_8) that are embedded within the two-stage SMART design, i.e.

Regime 1 (d_1): ('3','3'^{R},'4'^{NR})

Regime 2 (d_2): ('3','3'^{R},'5'^{NR})

Regime 3 (d_3): ('3','3'^{R},'6'^{NR})

Regime 4 (d_4): ('3','3'^{R},'7'^{NR})

Regime 5 (d_5): ('8','8'^{R},'4'^{NR})

Regime 6 (d_6): ('8','8'^{R},'5'^{NR})

Regime 7 (d_7): ('8','8'^{R},'6'^{NR})

Regime 8 (d_8): ('8','8'^{R},'7'^{NR})

Here, Regime 1 consists of treating a patient initially (after the basic treatment steps 1 and 2) with treatment 3, continuing with treatment 3 at the second stage if the patient is adjudged a responder (R) at

the end of stage 1, and switching to treatment 4 at the second stage if s/he is a non-responder (NR). All other regimes can be explained similarly.

There are a number of advantages of SMART designs over a series of single-stage trials to develop an optimal DRT (Chakraborty and Moodie, 2013). First, the single-stage trials may fail to detect delayed therapeutic effects. For example, the patients who receive laser therapy may achieve better short-term outcomes than those who receive SRP initially; however, the relative merit of SRP may be realized in the second stage when possible adverse events occur due to laser therapy. Second, SMARTs offer the option to reveal useful diagnostic information in that even though an initial treatment (e.g. SRP) may not be particularly effective, it can guide a better treatment choice (e.g., systemic anti-microbial therapy as an adjunct) at the second stage; a series of single stage trials fail to offer this diagnostic effect. Third, the non-responding patients to an initial treatment (e.g., SRP) may drop out of single-stage trials, but they are less likely to drop out of a SMART because they expect to receive possibly better treatments (e.g., SRP with adjuncts) at the next stage, even if SRP is not effective for them initially. Thus, the cohort of participants who are recruited and retained in a SMART may be quite different from those who are recruited and retained in a single-stage trial. This cohort effect may lead to biased estimation of effects of treatment sequences (DTRs) in single-stage trials, but should be well taken care of in a SMART.

3 SMART Design: Model, Hypothesis Testing and Sample Size Calculations

In this section, we propose the theoretical framework and an associated novel sample size formula for our SMART design.

3.1 Statistical Model

We start by introducing some notations. Let A_{i1} denote the treatment for patient i at the initial stage (i.e. '3' or '8'), $R_i(A_{i1})$ denote the proximal response after initial treatment A_{i1}, i.e., $R_i(\cdot) = 1$ if the i^{th} patient is a responder and $R_i(\cdot) = 0$ otherwise, $A_{i2}(A_{i1}, R_i(A_{i1}))$ denote the treatment at the final stage based on the initial treatment and proximal response, Y_{it} denote the final outcome measure, i.e. *change in mean CAL* for the t^{th} tooth of patient i, and M_{it} denote the missingness indicator of the t^{th} tooth of patient i, i.e., $M_{it} = 1$ if missing, or 0 otherwise. Our proposed model allows missing teeth to occur either before, or during the study, and the recorded Y_{it} at the end of the two-stage study is either a number (in mm), or missing (resulting from tooth t be missing before, or during the study). The expectation of the primary outcome Y_{it} considers the joint distribution of the mean CAL change and the missingness indicator, see Appendix C.

Thus, the observed data trajectory for patient i can be described as $\boldsymbol{O}_i$=(A_{i1}, $R_i(A_{i1})$, $A_{i2}(A_{i1}, R_i(A_{i1}))$, $Y_{i,1},\ldots,Y_{i,28}$, $M_{i,1},\ldots,M_{i,28}$). Note that there are N patients in the sample, and each patient has a maximum of 28 teeth (if no tooth is missing). Thus, the (overall) outcome measure for patient i is $\bar{Y}_i = \sum_{t=1}^{28} Y_{it}(1 - M_{it}) / \sum_{t=1}^{28}(1 - M_{it})$, which is the mean of CAL of the *available teeth* for patient i. Hence the proportion of available teeth for patient i is $\hat{p}_i = \sum_{t=1}^{28}(1 - M_{it})/28$. The generative model for $\boldsymbol{Y}$ is specified by a regression model of the form:

$$Y_{it} = \mu_i + Q_{it} + \epsilon_{it1}, \tag{1}$$

for $i = 1, \ldots, N$ and $t = 1, \ldots, 28$, where $\mu_i = \beta_0 + \beta_1 A_{i13} + \beta_2 A_{i13} R_i + \beta_3 R_i + \beta_4 A_{i13} A_{i24}(1 - R_i) + \beta_5 A_{i13} A_{i25}(1 - R_i) + \beta_6 A_{i13} A_{i26}(1 - R_i)$. Here, A_{i13} is an indicator of treatment '3' at initial stage for patient i, A_{i24} is an indicator of treatment '4' at final stage for patient i, and ϵ_{it1} is the (random) error term distributed as a skew-normal (SN), or skew-t (ST) density (Azzalini and Capitanio, 2003b), i.e., $\epsilon_{it1} \sim ST(0, \sigma_1^2, \lambda, \nu)$, with location parameter 0, scale parameter σ_1, skewness parameter λ, and degrees of freedom ν that measures the kurtosis. Note that the distribution of ϵ_{it1} is normal if $\lambda = 0$

and $\nu = \infty$, skew-normal if $\lambda \neq 0$ and $\nu = \infty$, t if $\lambda = 0$ and $\nu < \infty$, and skew-t, if $\lambda \neq 0$ and $\nu < \infty$. Expressions of the mean, variance, skewness γ_1 and kurtosis γ_2 for both SN and ST distributions are presented in Appendices A and B respectively. Following Reich and Bandyopadhyay (2010), we assume the latent vector $\boldsymbol{Q}_i$=$(Q_{i1}, \ldots, Q_{i28})^\top$ follows a multivariate normal distribution, with mean vector $\mathbf{0}_{28\times 1}$ and covariance matrix $\boldsymbol{\Sigma}_{28\times 28}$ with a conditional autoregressive (CAR) structure, i.e. $\boldsymbol{\Sigma}_{28\times 28} = \tau^2(\boldsymbol{C}_{28\times 28} - \rho \boldsymbol{D}_{28\times 28})^{-1}$. Here, $\tau^2 > 0$ and $\rho \in [0, 1]$ are the parameters controlling the magnitude of variation, and degree of spatial association, respectively. For the matrix $\boldsymbol{D}$, the elements $D_{tt'}$ are ones if locations t and t' are adjacent, and zeroes otherwise. The matrix $\boldsymbol{C}$ is diagonal with diagonal elements $C_{tt} = \sum_{t'} D_{tt'}$.

Next, under the assumption of non-randomly missing teeth (locations of missing teeth are not random, but rather related to the CP health in that region of the mouth), we propose a probit regression model for the missing teeth indicator as a function of the underlying (spatial) latent term $\boldsymbol{Q}_i$. Define $M_{it} = I(M_{it0} > 0)$, where M_{it0} is a (latent) continuous variable, modeled as:

$$M_{it0} = a_0 + b_0 Q_{it} + \epsilon_{it0}, \tag{2}$$

where $\epsilon_{it0} \overset{i.i.d}{\sim} N(0, \sigma_0^2)$. Note, the above probit trick allows us to connect the continuous (spatial) latent $\boldsymbol{Q}_i$ to the binary missingness indicator M_{it}. Thus, under a shared random parameter formulation (Albert, 2019) popularly used in longitudinal studies with informative (or missing not at random) dropout, and also in oral health studies (Reich and Bandyopadhyay, 2010), the latent variable $\boldsymbol{Q}_i$ not only models the association between teeth in a mouth, but also acts as the dependence term connecting the observed response and the missing data process. For the sake of identifiability, we choose $\sigma_0^2 = 1$. Here, under the popular shared-parameter framework (Vonesh *et al.*, 2006), $\boldsymbol{Q}$ facilitates sharing of information between $\boldsymbol{Y}$ and $\boldsymbol{M}$ for modeling non-randomly missing data. The parameters a_0, b_0, and the quantities $\boldsymbol{Q}_i$ and ϵ_{it0} determine the proportion of available teeth $p_i = E(\hat{p}_i)$, which can be estimated using either stochastic or deterministic method (see Appendix C). The parameter b_0 controls the association between $\boldsymbol{Y}$ and $\boldsymbol{M}$, e.g., $b_0 = 0$ indicates no association. Also, the Pearson correlation coefficient between Y_{it} and M_{it0} (from (2)) is c_{it} = $b_0 \text{var}(Q_{it}) / \sqrt{(\text{var}(Q_{it}) + \text{var}(\epsilon_{it1}))(b_0^2 \text{var}(Q_{it}) + \text{var}(\epsilon_{it0}))}$; see Appendix C for the derivation. For power analysis, one may choose the distributions of $\boldsymbol{Q}_i$, ϵ_{it1} and ϵ_{it0} from the literature, e.g., Reich and Bandyopadhyay (2010). Define c_i=$\sum_t c_{it}/28$. The clinician may suggest values for μ_i, p_i and c_i based on experience, or literature review, and the corresponding estimates of a_0 and b_0 can be obtained by solving a set of simultaneous equations involving p_i and c_i. However, when clinical advice is not available, sample size calculations should undergo a sensitivity analysis. In other words, estimates of N can be computed under various choices of the parameters, e.g., $p_i = 0.2, 0.3, 0.5 or 0.8$, and a conservative estimate of N maybe selected.

Next, we derive the expected value and variance for the sample mean of a DTR, using d_1 as an example, based on the IPW principle. IPW techniques have been successfully applied for estimating regression coefficients (Robins *et al.*, 1994b), and population mean (Cao *et al.*, 2009), in the context of incomplete data. For DTRs under SMART designs, most likely, we are unable to sample data directly from a particular regime. For example, responders of SRP can be viewed as coming from any of the regimes 1–4. Hence, a method of moments estimate of the sample mean for regime 1 is given by:

$$\bar{Y}^{d_1} = E_{d_1}(\bar{Y}_i) = E\left(W_i^{d_1} \bar{Y}_i\right), \tag{3}$$

where

$$W_i^{d_1} = \frac{I(A_{i1} = a_{1i}^{d_1}, A_{i2} = [a_{2i}^{d_1 R}]^{R_i} [a_{2i}^{d_1 NR}]^{1-R_i})}{\pi_{1i}^{d_1} [\pi_{2i}^{d_1 R}]^{R_i} [\pi_{2i}^{d_1 NR}]^{1-R_i}}. \tag{4}$$

where, $\bar{Y}_i$: the final outcome measure of CAL change for patient i; $I(\cdot)$: indicator function; R_i: binary response indicator for treatment '3' at initial stage; $a_{1i}^{d_1}$: the regime 1 (d_1) treatment at initial stage for

patient i, e.g., '3'; $a_{2i}^{d_1 R}$: regime 1 treatment at final stage if participant i is a responder (i.e. $R_i = 1$), e.g., '3'; $a_{2i}^{d_1 NR}$: regime 1 treatment at final stage if patient i is a non-responder (i.e. $R_i = 0$), e.g., '4'; $\pi_{1i}^{d_1}$: probability of treatment allocation of regime 1 at initial stage for patient i; $\pi_{2i}^{d_1 R}$: probability of treatment allocation of regime 1 at final (2nd) stage if patient i is a responder, i.e. 1; $\pi_{2i}^{d_1 NR}$: probability of treatment allocation of regime 1 at final stage, if patient i is a non-responder, i.e. 1/4.

To maximize power, we estimate $\pi_{1i}^{d_1}$ as in Murphy (2005) to have equal sample sizes across all possible regimes. Based on Figure 2, we set

$$\pi_{1i}^{d_1} = \frac{(1 \cdot \gamma^{d_1} + \frac{1}{4} \cdot (1 - \gamma^{d_1}))^{-1}}{(1 \cdot \gamma^{d_1} + \frac{1}{4} \cdot (1 - \gamma^{d_1}))^{-1} + (1 \cdot \gamma^{d_5} + \frac{1}{4} \cdot (1 - \gamma^{d_5}))^{-1}}, \tag{5}$$

where γ^{d_1} denotes the response rate for regime 1 at initial stage. When both γ^{d_1} and γ^{d_5} are unknown, Murphy (2005) suggested setting $\pi_{1i}^{d_1}$ to

$$\pi_{1i}^{d_1} = \frac{\max(1^{-1}, \frac{1}{4}^{-1})}{\max(1^{-1}, \frac{1}{4}^{-1}) + \max(1^{-1}, \frac{1}{4}^{-1})}. \tag{6}$$

Alternatively, we can also set $\pi_{1i}^{d_1} = 1/2$, ensuring equal probability of treatment allocation at the initial stage.

The mean and variance of $\bar{Y}^{d_1}$ are given by

$$E(\bar{Y}^{d_1}) = E(\frac{I(A_{i1} = a_{1i}^{d_1}, A_{i2} = [a_{2i}^{d_1 R}]^{R_i}[a_{2i}^{d_1 NR}]^{1-R_i})}{\pi_{1i}^{d_1}[\pi_{2i}^{d_1 R}]^{R_i}[\pi_{2i}^{d_1 NR}]^{1-R_i}} \bar{Y}_i), \tag{7}$$

and

$$\text{var}(\bar{Y}^{d_1}) = \frac{1}{N}\text{var}(\frac{I(A_{i1} = a_{1i}^{d_1}, A_{i2} = [a_{2i}^{d_1 R}]^{R_i}[a_{2i}^{d_1 NR}]^{1-R_i})}{\pi_{1i}^{d_1}[\pi_{2i}^{d_1 R}]^{R_i}[\pi_{2i}^{d_1 NR}]^{1-R_i}} \bar{Y}_i). \tag{8}$$

In terms of $\boldsymbol{\pi}$, $\boldsymbol{\gamma}$, $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$, (7) and (8) can be expressed alternatively as

$$E(\bar{Y}^{d_1}) = \gamma^{d_1}\mu_{d_1 R} + (1 - \gamma^{d_1})\mu_{d_1 NR} = \mu_{d_1}$$

and

$$\begin{aligned} & V(\bar{Y}^{d_1}) \\ = & \frac{1}{N}\{\frac{\gamma^{d_1}}{\pi_{1i}^{d_1}[\pi_{2i}^{d_1 R}]}(\sigma_{d_1 R}^2 + (1 - \pi_{1i}^{d_1}[\pi_{2i}^{d_1 R}])\mu_{d_1 R}^2) + \\ & \frac{1 - \gamma^{d_1}}{\pi_{1i}^{d_1}[\pi_{2i}^{d_1 NR}]}(\sigma_{d_1 NR}^2 + (1 - \pi_{1i}^{d_1}[\pi_{2i}^{d_1 NR}])\mu_{d_1 NR}^2) + \\ & \gamma^{d_1}(1 - \gamma^{d_1})(\mu_{d_1 R} - \mu_{d_1 NR})^2\}, \end{aligned}$$

where $\sigma_{d_1 R}^2$: the variance of $\bar{Y}_i$ from d_1 and $R_i = 1$; $\mu_{d_1 R}$: the mean of $\bar{Y}_i$ from d_1 and $R_i = 1$. Detailed expressions of both $\sigma_{d_1 R}^2$ and $\mu_{d_1 R}$ appear in Appendix C. However, we compute them using Monte Carlo method. Likewise, both $E(\bar{Y}^{d_3})$ and $V(\bar{Y}^{d_3})$ can be derived; see Appendix C.

3.2 Hypothesis Testing

Our SMART design allows the following important hypothesis tests:

1. Detecting a single DTR effect on CAL, using $H_0: \mu_{d_1} = 0$ vs $H_1: \mu_{d_1} = \delta_{d_1} \neq 0$;
2. Comparing two DTRs that share an initial treatment (e.g., DTRs starting with SRP as initial treatment and then followed by various adjuncts), using $H_0: \mu_{d_1} - \mu_{d_3} = 0$ vs $H_1: \mu_{d_1} - \mu_{d_3} = \delta_{d_1-d_3} \neq 0$;
3. Comparing two DTRs that do not share an initial treatment (e.g., DTRs initiated by SRP and laser respectively), using $H_0: \mu_{d_1} - \mu_{d_5} = 0$ vs $H_1: \mu_{d_1} - \mu_{d_5} = \delta_{d_1-d_5} \neq 0$.
4. Detecting one regime (e.g., regime 1) is the best among all the possible eight embedded regimes in Figure 2, i.e.

   $$H_0: \mu_{d_1} - \mu_{d_2} \leq 0 \text{ or } \mu_{d_1} - \mu_{d_3} \leq 0 \text{ or} \cdots \text{or } \mu_{d_1} - \mu_{d_8} \leq 0$$

   versus

   $$H_1: \mu_{d_1} - \mu_{d_2} > 0 \text{ and } \mu_{d_1} - \mu_{d_3} > 0 \text{ and} \cdots \text{and } \mu_{d_1} - \mu_{d_8} > 0.$$

Hypothesis 1 can be used to test whether the improvement in CAL in the proposed DTR is better than SRP (e.g., $\geq$ 0.5mm), or not worse than SRP with adjuncts (e.g., 0.7–1.1 mm), based on the systematic review results of Smiley *et al.* (2015). Since the network meta-analyses by John *et al.* (2017) found no significant evidence of CAL improvement among adjuncts, Hypothesis 2 can be used to test if indeed there are statistically significant differences between the DTRs of SRP and 'SRP + adjuncts'. The treatment effect of laser therapy is still under investigation; we can use Hypothesis 3 to test if there is a statistically significant difference between the DTRs initiated by SRP and laser. Finally, Hypothesis 4 can assist us in testing the finding by John *et al.* (2017) that SRP with local antimicrobial therapy is possibly the best when compared to SRP with other adjuncts.

Consider Hypothesis 2. The Expectation and variance of the difference in regime means can be expressed respectively as

$$E(\bar{Y}^{d_1} - \bar{Y}^{d_3}) = \mu_{d_1} - \mu_{d_3} = \delta_{d_1-d_3} \tag{9}$$

and

$$V(\bar{Y}^{d_1} - \bar{Y}^{d_3}) = V(\bar{Y}^{d_1}) + V(\bar{Y}^{d_3}) - 2COV(\bar{Y}^{d_1}, \bar{Y}^{d_3}) = \frac{1}{N} 2\sigma^2_{d_1-d_3}. \tag{10}$$

Note, both δ and σ^2 in equations (9) and (10) respectively are functions of the parameter vector $\mathbf{\Omega}_{d_1-d_3}$=($\mu$, τ, ρ, λ, ν, σ_1^2, σ_0^2, a_0, b_0, γ^{d_1}, $\pi_1^{d_1}$, $\pi_2^{d_1R}$, $\pi_2^{d_1NR}$, γ^{d_3}, $\pi_1^{d_3}$, $\pi_2^{d_3R}$, $\pi_2^{d_3NR}$). Also, μ, τ, ρ, λ, ν, σ_1^2, σ_0^2, a_0 and b_0 are defined in equations (1) and (2), while parameters γs and πs are defined in equations (3) to (5). The covariance between $\bar{Y}^{d_1}$ and $\bar{Y}^{d_3}$ is

$$\begin{aligned}
&\text{COV}\left(\bar{Y}^{d_1}, \bar{Y}^{d_3}\right) \\
=&\frac{1}{N}\{\frac{\gamma^{d_1}}{\pi_{1i}^{d_1}\pi_{2i}^{d_1R}}(\sigma^2_{d_1R} + \mu^2_{d_1R}) - \gamma^{d_1}\gamma^{d_3}\mu_{d_1R}\mu_{d_3R} \\
&- \gamma^{d_1}(1-\gamma^{d_3})\mu_{d_1R}\mu_{d_3NR} \\
&- \gamma^{d_3}(1-\gamma^{d_1})\mu_{d_1NR}\mu_{d_3R} \\
&- (1-\gamma^{d_1})(1-\gamma^{d_3})\mu_{d_1NR}\mu_{d_3NR}\}.
\end{aligned}$$

Note that $\gamma^{d_1} = \gamma^{d_3}$, $\mu_{d_1R} = \mu_{d_3R}$ and $\sigma^2_{d_1R} = \sigma^2_{d_3R}$, since the responders from treatment '3' are consistent with both the regimes 1 and 3. The derivations of both $E(\bar{Y}^{d_1} - \bar{Y}^{d_3})$ and $V(\bar{Y}^{d_1} - \bar{Y}^{d_3})$ can be found in Appendix C. Below we present a theoretical result following a set of assumptions; see Appendix D for the proof.

Assumptions:

1. Random vectors $(\bar{Y}_i, W_i^{d_1}, W_i^{d_3})$, $0 \leq i \leq N$ are independent and identically distributed, and distribution of $\bar{Y}_i$ is independent of $W_i^{d_1}$ and $W_i^{d_3}$, where $W_i^{d_1}$ is defined by equation (4).
2. Let Θ, the parameter space for the regime means under consideration (e.g., μ_{d_1} and μ_{d_3}), be a compact subset of real numbers. Also let μ_{d_10}, μ_{d_30}, and $\delta_{(d_1-d_3)0}$ denote the true values of μ_{d_1}, μ_{d_3}, and $\delta_{d_1-d_3}$, respectively. Assume unbiased estimators $\frac{1}{N_2}\sum_{i=1}^{N_2} W_i^{d_1}\bar{Y}_i$ and $\frac{1}{N_2}\sum_{i=1}^{N_2} W_i^{d_3}\bar{Y}_i$ for μ_{d_1} and μ_{d_3}, respectively, i.e., $E(W_i^{d_1}\bar{Y}_i - \mu_{d_1}) = 0$ only when $\mu_{d_1} = \mu_{d_10}$ and $E(W_i^{d_3}\bar{Y}_i - \mu_{d_3}) = 0$ only when $\mu_{d_3} = \mu_{d_30}$. Hence, $E(\hat{\delta}_{d_1-d_3} - \delta_{d_1-d_3}) = 0$ only for $\delta_{d_1-d_3} = \delta_{(d_1-d_3)0}$=$\mu_{d_10} - \mu_{d_30}$. Note, N_2 represents sample size N of Hypothesis 2.

Theorem 3.1 *The IPW and MOM estimator $\hat{\delta}_{d_1-d_3}$ is a consistent estimator of $\delta_{d_1-d_3}$. Under moment conditions and the above assumptions, we have $\sqrt{N_2}(\hat{\delta}_{d_1-d_3} - \delta_{(d_1-d_3)0}) \to Normal(0, 2\sigma^2_{d_1-d_3})$.*

Theorem 3.1 is based on the Hypothesis 2. However, we can also derive similar theorems for other hypothesis. Though the regimes 1 and 5 do not share any initial treatments, the covariance between the sample mean of these two regimes can be derived in the similar way as COV $\left(\bar{Y}^{d_1}, \bar{Y}^{d_3}\right)$. The mathematical formula for COV $\left(\bar{Y}^{d_1}, \bar{Y}^{d_5}\right)$ is given in Appendix C.

Before deriving the sample size formula, we present the test statistics for the corresponding hypotheses below. For example, for $H_0 : \mu_{d_1} - \mu_{d_3} = 0$ vs. $H_1 : \mu_{d_1} - \mu_{d_3} = \delta_{d_1-d_3} \neq 0$ (hypothesis 2), we use the univariate Wald statistic $Z_2 = \hat{\delta}_{d_1-d_3}/\sqrt{2\sigma^2_{d_1-d_3}/N_2}$, where $\hat{\delta}_{d_1-d_3}$=$\bar{Y}^{d_1} - \bar{Y}^{d_3}$ is the estimated effect size and $\sigma^2_{d_1-d_3}$ is given in (10). In large samples, Z_2 follows a standard normal distribution if H_0 is true. Hence, at α level of significance, we reject H_0 if $\mid Z_2 \mid > z_{\alpha/2}$, where $z_{\alpha/2}$ is the upper $\alpha/2$ quantile of a standard normal distribution. In a similar way, the test statistic for hypothesis 1 ($H_0 : \mu_{d_1} = 0$ vs. $H_1 : \mu_{d_1} = \delta_{d_1} \neq 0$) and hypothesis 3 ($H_0 : \mu_{d_1} - \mu_{d_5} = 0$ vs $H_1 : \mu_{d_1} - \mu_{d_5} = \delta_{d_1-d_5} \neq 0$) are $Z_1 = \hat{\delta}_{d_1}/\sqrt{2\sigma^2_{d_1}/N_1}$ and $Z_3 = \hat{\delta}_{d_1-d_5}/\sqrt{2\sigma^2_{d_1-d_5}/N_3}$ respectively, where both follow standard normal distribution under H_0.

Note, Hypothesis 4 uses a multivariate test statistic $\boldsymbol{Z}$=$(Z_{d_1-d_2}, Z_{d_1-d_3}, \ldots, Z_{d_1-d_8})^\top$, where each component of the vector, e.g., $Z_{d_1-d_2}$=$\hat{\delta}_{d_1-d_2}/\sqrt{2\sigma^2_{d_1-d_2}/N_4}$, follows a standard normal distribution, under H_0. Hence, at α level of significance, we reject H_0 if all of $Z_{d_1-d_2}, \ldots, Z_{d_1-d_8} > z_\alpha$. The first three hypothesis tests are two-sided tests, while the fourth one is one-sided. Thus, in order to compute the joint probability Pr($Z_{d_1-d_2} > z_\alpha, \ldots, Z_{d_1-d_8} > z_\alpha$), one requires the covariance matrix of $(\hat{\delta}_{d_1-d_2}, \ldots, \hat{\delta}_{d_1-d_8})^\top$; the derivation of COV($\hat{\delta}_{d_1-d_2}, \hat{\delta}_{d_1-d_3}$) is presented in Appendix C.

3.3 Sample Size Calculation

The calculated sample size is capable to detect the postulated effect size corresponding to either a single regime, or the difference between two regimes, or, if one regime is superior to the others. The proposed sample size formulas for Hypothesis tests 1-4 under our SMART design are given by

$$N_1 = 2(z_{\alpha/2} - z_{1-\beta})^2 \frac{\sigma^2_{d_1}}{\delta^2_{d_1}}, \tag{11}$$

$$N_2 = 2(z_{\alpha/2} - z_{1-\beta})^2 \frac{\sigma^2_{d_1-d_3}}{\delta^2_{d_1-d_3}}, \tag{12}$$

$$N_3 = 2(z_{\alpha/2} - z_{1-\beta})^2 \frac{\sigma^2_{d_1-d_5}}{\delta^2_{d_1-d_5}} \tag{13}$$

while N_4 is the root of the equation

$$\Pr\left\{Z_{d_1-d_2} \leq -(z_\alpha - \delta^*_{d_1-d_2}\sqrt{N_4/2}), \ldots, Z_{d_1-d_8} \leq -(z_\alpha - \delta^*_{d_1-d_8}\sqrt{N_4/2})\right\} = \text{power}, \tag{14}$$

respectively, where $\sigma^2_{d_1-d_3}$ is defined by (10), and similarly, both $\sigma^2_{d_1}$ and $\sigma^2_{d_1-d_5}$ can also be defined; α =Pr(Type I error), β =Pr(Type II error)= 1 - Power, $\Pr(z > z_{\alpha/2}) = \alpha/2$ and $\Pr(z > z_{1-\beta}) = 1-\beta$, the effect size $\delta_{d_1-d_3} = \mu_{d_1} - \mu_{d_3}$; $Z_{d_1-d_2}$,..., $Z_{d_1-d_8}$ have standard normal distributions and same correlation matrix as $\hat{\delta}_{d_1-d_2}$, ..., $\hat{\delta}_{d_1-d_8}$; see the derivation in Appendix C. Therefore, we define the standardized effect size by $\delta^*_{d_1-d_3} = \delta_{d_1-d_3}/\sigma_{d_1-d_3}$. Note, our calculations advance the previous ones for SMART designs in clustered data (Ghosh *et al.*, 2016; NeCamp *et al.*, 2017) by including non-Gaussianity, spatial association, and non-random missingness features, typical of periodontal responses, in addition to considering comparisons between regimes that share the same initial treatment, and even testing whether one particular regime is the best among all possible embedded regimes. Also, the subjects (or clusters) in our setting are randomly allocated with equal probability for each embedded regime, which was suggested by Murphy (2005) to maximize the power for comparing regime means.

4 Simulation Studies

We now present simulation studies to investigate the finite-sample performance of the proposed sample size formulas (11) to (14) in terms of computing Monte Carlo power estimates based on 5000 simulated data sets, given the Type-II error rate $\beta = 0.2$, or nominal power of 80% and Type-I error rate $\alpha = 0.05$ (hypotheses 1–3) or 0.025 (hypothesis 4). The Monte Carlo data generation steps are given below. These include generating the random variables A_{i1}, $R_i(A_{i1})$ and $A_{i2}(A_{i1}, R_i(A_{i1}))$, M_{it} and Y_{it} for each patient i, $i = 1, \ldots, N$.

Step 1. The initial treatment A_{i1} is assigned randomly to either '3' or '8', with probability $\pi^{d_1}_{1i}$ and $1 - \pi^{d_1}_{1i}$ respectively.

Step 2. The response variable $R_i(A_{i1})$ is generated from Bernoulli(γ^{d_1}) if A_{i1}='3', or Bernoulli(γ^{d_5}) if A_{i1}='8', where $\gamma^{d_1} = 0.25$ or 0.5 and $\gamma^{d_5} = 0.5$.

Step 3. The final treatment $A_{i2}(A_{i1} = \text{'3'}, R_i(A_{i1} = \text{'3'}) = 1)$ is assigned to '3' with probability of 1, and $A_{i2}(A_{i1} = \text{'3'}, R_i(A_{i1} = \text{'3'}) = 0)$ is randomly assigned to'4', '5', '6' or '7' with probability of 1/4, while $A_{i2}(A_{i1} = \text{'8'}, R_i(A_{i1} = \text{'8'}) = 1)$ is assigned to '8' with probability 1 and $A_{i2}(A_{i1} = \text{'8'}, R_i(A_{i1} = \text{'8'}) = 0)$ is assigned to '4', '5', '6' or '7' with probability of 1/4.

Step 4. The change in mean CAL Y_{it} and missingness indicator M_{it} of each tooth are generated by regression models (1) and (2) respectively. For the model parameters, we assume $\tau = 0.85$, $\rho = 0.975$, $a_0 = -1$, $b_0 = 0.5$, $\sigma_1 = 0.95$ and $\sigma_0 = 1$, based on estimates from Reich and Bandyopadhyay (2010). For model (1), we select $\mu_i = 0$ if $R_i = 1$ and μ_i = 0.5, 2 or 5 if $R_i = 0$ to test the proposed method. The skewness and kurtosis parameters for the error term ϵ_{it1} of model (1) are chosen as $\lambda = 0$, 2 and 10, $\nu = \infty$, 8, 5 and 3. This choice of parameters for model (2) give expected proportion of

available teeth around 80% (i.e. $p_i \approx 0.8$) for each patient. Given the parameters of models (1) and (2), at $\lambda = 10$ and $\nu = 3$, the association between CAL change and missingness is around 0.42 (i.e. $c_i \approx 0.42$).

Step 5. The mean CAL change for patient i is computed as $\bar{Y}_i = \frac{\sum_{t=1}^{28} Y_{it}(1 - M_{it})}{\sum_{t=1}^{28}(1 - M_{it})}$.

Tables 1 - 4 present a list of sample sizes calculated by the proposed method, with the corresponding estimated Monte Carlo powers based on hypotheses 1–4, respectively. Various pairs of the skewness and kurtosis parameters (λ, ν) corresponding to the error ϵ_{it1} are considered. Recall, $\lambda = 0$ and $\nu = \infty$ indicates normal distribution; $\lambda \neq 0$ and $\nu = \infty$ indicates skew-normal distribution; $\lambda = 0$ and $\nu < \infty$ indicates t distribution, and $\lambda \neq 0$ and $\nu < \infty$ indicates skew-t distribution. We consider a list of effect sizes (e.g., δ_{d_1}=μ_{d_1}), and present their corresponding absolute values (e.g., $\mid \delta_{d_1} \mid$), and the absolute value of the standardized effect size, i.e. $\mid \delta^*_{d_1} \mid$=$\mid \delta_{d_1}/\sigma_{d_1} \mid$ except for Table 4, where we present a list of average of absolute values of effect sizes (e.g., $\overline{\mid \delta \mid}$=$\sum_{j=2}^{8} \mid \delta_{d_1-d_j} \mid /7$), and the corresponding standardized effect size (i.e. $\overline{\mid \delta^* \mid} = \sum_{j=2}^{8} \mid \delta_{d_1-d_j}/\sigma_{d_1-d_j} \mid /7$).

Table 1 detects a single regime (i.e. regime 1) effect. We define $\mu_i = 0$ if $R_i = 1$ and $\mu_i = 2$ (e.g., the first row), or 5 (e.g., the second row) if $R_i = 0$. Table 2 compares two regimes that share an initial treatment (i.e. regimes 1 versus 3). We define $\mu_i = 0$ if $R_i = 1$ and $\mu_i = 0.5$ if $R_i = 0$ for regime 1, while we define $\mu_i = 0$ if $R_i = 1$ and $\mu_i = 2$ (e.g., the first row) or 5 (e.g., the second row) if $R_i = 0$ for regime 3. Table 3 compares two regimes without sharing initial treatment (i.e. regimes 1 versus 5), where we define $\mu_i = 0$ if $R_i = 1$ and $\mu_i = 0.5$ if $R_i = 0$ for regime 1, while we define $\mu_i = 0$ if $R_i = 1$ and $\mu_i = 2$ (e.g., the first row) or 5 (e.g., the second row) if $R_i = 0$ for regime 5. Finally, Table 4 detects if regime one is better than rest of the embedded regimes (i.e. regimes 1 versus 2–8), where we define $\mu_i = 0$ if $R_i = 1$ and $\mu_i = 2$ (e.g., the first row) or 5 (e.g., the second row) if $R_i = 0$ for regime 1, while we define $\mu_i = 0$ if $R_i = 1$ and $\mu_i = 0$ if $R_i = 0$ for regimes 2–8.

The results are summarized below. We obtain mostly small to medium (0.2–0.5) absolute standardized effect sizes, except some smaller effects (< 0.2) in Table 3. The tables show that the Monte Carlo estimated powers ($78\% - 82\%$) are close to the nominal power based on the sample size formulas (11)-(14). Note, the sample sizes are rounded to the next integer. Throughout our simulation studies, we assume $\tau = 0.85$ and $\rho = 0.975$, based on the literature (Reich and Bandyopadhyay, 2010). Sensitivity checks (omitted here for brevity) revealed the estimated N to increase with increase of either τ, or ρ, or both. With respect to calibration of ρ as a (spatial) correlation measure, a very high value of rho (say, $\rho = 0.99$) only translates to Moran's$I \leq 0.5$, indicative of moderate correlation (Banerjee *et al.*, 2014). Hence, under the assumption of the presence of meaningful spatial association, we suggest to input a very high value of ρ during sample size calculations.

Table 1 Formula-based sample size (N_1) and estimated Monte Carlo power ($\hat{P}$) at $\beta = 0.2$ and $\alpha = 0.05$, based on hypothesis test of $H_0: \mu_{d_1} = 0$ versus $H_1: \mu_{d_1} \neq 0$, under varying absolute effect size ($\mid \delta_{d_1} \mid$) or standardized effect size ($\mid \delta^*_{d_1} \mid$), treatment '3' response rate (γ^{d_1}), skewness (λ) and degrees of freedom (ν), given treatment '8' response rate $\gamma^{d_5} = 0.5$, $\sigma_1 = 0.95$, $\sigma_0 = 1$, $\rho = 0.975$, $\tau = 0.85$, expected % of available teeth per patient $p_i = 80\%$.

Absolute Effect Size $\mid \delta_{d_1} \mid$	Skewness λ	Degrees of Freedom ν	Response Rate γ^{d_1}	Standardized Effect Size $\mid \delta^*_{d_1} \mid$	Formula-based Sample Size N_1	Monte Carlo Power $\hat{P}$
1.28	0	Inf	0.25	0.43	84.00	0.80
3.53				0.48	68.00	0.79
0.78			0.5	0.29	189.00	0.79
2.28				0.34	135.00	0.80
1.28		8	0.25	0.43	84.00	0.80
3.53				0.48	68.00	0.79
0.78			0.5	0.29	190.00	0.81
2.28				0.34	135.00	0.80
1.27		5	0.25	0.43	85.00	0.80
3.53				0.48	68.00	0.80
0.78			0.5	0.29	192.00	0.79
2.28				0.34	135.00	0.80
1.28		3	0.25	0.43	86.00	0.81
3.52				0.48	68.00	0.80
0.78			0.5	0.28	196.00	0.79
2.28				0.34	135.00	0.79
1.96	2	Inf	0.25	0.51	61.00	0.79
4.20				0.51	61.00	0.81
1.45			0.5	0.41	92.00	0.80
2.95				0.39	102.00	0.78
2.02		8	0.25	0.51	60.00	0.79
4.27				0.51	61.00	0.81
1.53			0.5	0.42	88.00	0.80
3.03				0.40	100.00	0.79
2.09		5	0.25	0.52	59.00	0.80
4.33				0.51	60.00	0.80
1.58			0.5	0.43	85.00	0.79
3.08				0.40	98.00	0.79
2.22		3	0.25	0.52	58.00	0.80
4.46				0.52	60.00	0.79
1.71			0.5	0.44	80.00	0.79
3.21				0.41	94.00	0.79
2.03	10	Inf	0.25	0.51	60.00	0.79
4.28				0.51	61.00	0.81
1.53			0.5	0.43	87.00	0.80
3.03				0.40	99.00	0.78
2.11		8	0.25	0.52	59.00	0.79
4.36				0.51	60.00	0.80
1.61			0.5	0.44	83.00	0.79
3.11				0.40	97.00	0.79
2.17		5	0.25	0.52	58.00	0.79
4.42				0.52	60.00	0.80
1.67			0.5	0.44	81.00	0.80
3.18				0.41	95.00	0.79
2.31		3	0.25	0.53	57.00	0.80
4.57				0.52	59.00	0.79
1.82			0.5	0.46	76.00	0.80
3.32				0.42	92.00	0.80

Table 2 Formula-based sample size (N_2) and estimated Monte Carlo power ($\hat{P}$) at $\beta = 0.2$ and $\alpha = 0.05$, based on hypothesis test of $H_0 : \mu_{d_1} = \mu_{d_3}$ versus $H_1 : \mu_{d_1} \neq \mu_{d_3}$, under varying absolute effect size ($\mid \delta_{d_1-d_3} \mid$) or standardized effect size ($\mid \delta^*_{d_1-d_3} \mid$) and treatment '3' response rate (γ^{d_1}), skewness (λ) and degrees of freedom (ν), given treatment '8' response rate $\gamma^{d_5} = 0.5$, $\sigma_1 = 0.95$, $\sigma_0 = 1$, $\rho = 0.975$, $\tau = 0.85$, expected $\%$ of available teeth per patient $p_i = 80\%$.

Absolute Effect Size $\mid \delta_{d_1-d_3} \mid$	Skewness λ	Degrees of Freedom ν	Response Rate γ^{d_1}	Standardized Effect Size $\mid \delta^*_{d_1-d_3} \mid$	Formula-based Sample Size N_2	Monte Carlo Power $\hat{P}$
1.12	0	Inf	0.25	0.35	127.00	0.80
3.37				0.45	77.00	0.81
0.75			0.5	0.26	229.00	0.79
2.25				0.33	141.00	0.80
1.12		8	0.25	0.35	128.00	0.79
3.38				0.45	77.00	0.79
0.75			0.5	0.26	233.00	0.80
2.25				0.33	141.00	0.80
1.13		5	0.25	0.35	128.00	0.79
3.37				0.45	77.00	0.79
0.75			0.5	0.26	234.00	0.80
2.25				0.33	141.00	0.79
1.12		3	0.25	0.35	132.00	0.80
3.37				0.45	77.00	0.79
0.75			0.5	0.26	238.00	0.80
2.25				0.33	142.00	0.80
1.13	2	Inf	0.25	0.25	242.00	0.80
3.38				0.39	104.00	0.79
0.75			0.5	0.19	435.00	0.80
2.25				0.29	188.00	0.81
1.13		8	0.25	0.25	257.00	0.79
3.37				0.38	107.00	0.79
0.75			0.5	0.18	461.00	0.80
2.25				0.28	194.00	0.80
1.12		5	0.25	0.24	275.00	0.80
3.37				0.38	110.00	0.79
0.75			0.5	0.18	486.00	0.80
2.25				0.28	199.00	0.80
1.13		3	0.25	0.23	305.00	0.79
3.38				0.37	116.00	0.79
0.75			0.5	0.17	551.00	0.80
2.25				0.27	211.00	0.80
1.13	10	Inf	0.25	0.25	257.00	0.80
3.38				0.38	107.00	0.80
0.75			0.5	0.18	465.00	0.81
2.25				0.28	195.00	0.80
1.13		8	0.25	0.24	277.00	0.80
3.37				0.38	111.00	0.80
0.75			0.5	0.18	498.00	0.80
2.25				0.28	201.00	0.79
1.12		5	0.25	0.23	293.00	0.80
3.37				0.37	114.00	0.79
0.75			0.5	0.17	527.00	0.80
2.25				0.28	206.00	0.79
1.12		3	0.25	0.22	333.00	0.79
3.38				0.36	121.00	0.79
0.75			0.5	0.16	593.00	0.80
2.25				0.27	220.00	0.80

Table 3 Formula-based sample size (N_3) and estimated Monte Carlo power ($\hat{P}$) at $\beta = 0.2$ and $\alpha = 0.05$, based on hypothesis test of $H_0: \mu_{d_1} = \mu_{d_5}$ versus $H_1: \mu_{d_1} \neq \mu_{d_5}$, under varying absolute effect size ($\mid \delta_{d_1-d_5} \mid$) or standardized effect size ($\mid \delta^*_{d_1-d_5} \mid$) and treatment '3' response rate (γ^{d_1}), skewness (λ) and degrees of freedom (ν), given treatment '8' response rate $\gamma^{d_5} = 0.5$, $\sigma_1 = 0.95$, $\sigma_0 = 1$, $\rho = 0.975$, $\tau = 0.85$, expected $\%$ of available teeth per patient $p_i = 80\%$.

Absolute Effect Size $\mid \delta_{d_1-d_5} \mid$	Skewness λ	Degrees of Freedom ν	Response Rate γ^{d_1}	Standardized Effect Size $\mid \delta^*_{d_1-d_5} \mid$	Formula-based Sample Size N_3	Monte Carlo Power $\hat{P}$
0.63	0	Inf	0.25	0.19	420.00	0.78
2.12				0.28	196.00	0.79
0.75			0.5	0.25	244.00	0.80
2.25				0.33	142.00	0.79
0.62		8	0.25	0.19	432.00	0.81
2.13				0.28	196.00	0.79
0.74			0.5	0.25	249.00	0.80
2.25				0.33	143.00	0.80
0.62		5	0.25	0.19	435.00	0.79
2.12				0.28	197.00	0.81
0.75			0.5	0.25	248.00	0.80
2.25				0.33	143.00	0.80
0.63		3	0.25	0.19	444.00	0.80
2.12				0.28	198.00	0.80
0.75			0.5	0.25	257.00	0.80
2.25				0.33	143.00	0.79
0.62	2	Inf	0.25	0.14	792.00	0.79
2.12				0.24	263.00	0.79
0.75			0.5	0.19	452.00	0.80
2.25				0.29	190.00	0.80
0.62		8	0.25	0.14	847.00	0.79
2.13				0.24	270.00	0.80
0.75			0.5	0.18	486.00	0.80
2.25				0.28	198.00	0.80
0.63		5	0.25	0.13	886.00	0.79
2.13				0.24	277.00	0.81
0.75			0.5	0.18	508.00	0.80
2.25				0.28	202.00	0.80
0.63		3	0.25	0.13	1000.00	0.80
2.13				0.23	293.00	0.79
0.75			0.5	0.16	580.00	0.80
2.25				0.27	215.00	0.80
0.63	10	Inf	0.25	0.14	842.00	0.79
2.12				0.24	271.00	0.80
0.75			0.5	0.18	484.00	0.81
2.25				0.28	196.00	0.80
0.63		8	0.25	0.13	909.00	0.80
2.12				0.24	281.00	0.80
0.75			0.5	0.17	524.00	0.81
2.25				0.28	204.00	0.80
0.63		5	0.25	0.13	960.00	0.79
2.13				0.23	286.00	0.81
0.75			0.5	0.17	550.00	0.80
2.25				0.27	210.00	0.81
0.62		3	0.25	0.12	1097.00	0.80
2.12				0.23	307.00	0.80
0.75			0.5	0.16	623.00	0.80
2.25				0.27	224.00	0.80

Table 4 Formula-based sample size (N_4) and estimated Monte Carlo power ($\hat{P}$) at $\beta = 0.2$ and $\alpha = 0.025$, based on hypothesis test of $H_0: \mu_{d_1} \leq \mu_{d_2}$ or ... or $\mu_{d_1} \leq \mu_{d_8}$ versus $H_1: \mu_{d_1} > \mu_{d_2}$ & ... & $\mu_{d_1} > \mu_{d_8}$, under varying average of absolute effect sizes ($\overline{\mid \delta \mid} = \sum_{j=2}^{8} \mid \delta_{d_1-d_j} \mid /7$) or standardized effect size ($\overline{\mid \delta^* \mid} = \sum_{j=2}^{8} \mid \delta_{d_1-d_j}/\sigma_{d_1-d_j} \mid /7$) and treatment '3' response rate (γ^{d_1}), skewness (λ) and degrees of freedom (ν), given treatment '8' response rate $\gamma^{d_5} = 0.5$, $\sigma_1 = 0.95$, $\sigma_0 = 1$, $\rho = 0.975$, $\tau = 0.85$, expected % of available teeth per patient $p_i = 80\%$.

Absolute Effect Size $\mid \delta \mid$	Skewness λ	Degrees of Freedom ν	Response Rate γ^{d_1}	Standardized Effect Size $\mid \delta^* \mid$	Formula-based Sample Size N_4	Monte Carlo Power $\hat{P}$
1.50	0	Inf	0.25	0.48	95.00	0.79
3.75				0.51	71.00	0.80
1.00			0.5	0.35	179.00	0.80
2.50				0.37	132.00	0.81
1.50		8	0.25	0.48	96.00	0.80
3.75				0.51	71.00	0.81
1.00			0.5	0.35	180.00	0.80
2.50				0.37	132.00	0.80
1.50		5	0.25	0.47	97.00	0.80
3.75				0.51	71.00	0.80
1.00			0.5	0.35	182.00	0.81
2.50				0.37	132.00	0.79
1.50		3	0.25	0.47	101.00	0.80
3.75				0.51	72.00	0.81
1.00			0.5	0.34	189.00	0.80
2.50				0.37	134.00	0.79
1.50	2	Inf	0.25	0.36	161.00	0.79
3.75				0.44	93.00	0.79
1.00			0.5	0.26	297.00	0.79
2.50				0.33	172.00	0.79
1.50		8	0.25	0.35	173.00	0.78
3.75				0.43	97.00	0.79
1.00			0.5	0.26	318.00	0.78
2.50				0.32	179.00	0.79
1.50		5	0.25	0.34	183.00	0.78
3.75				0.43	100.00	0.79
1.00			0.5	0.25	336.00	0.78
2.50				0.32	184.00	0.79
1.50		3	0.25	0.32	209.00	0.80
3.75				0.42	107.00	0.79
1.00			0.5	0.24	383.00	0.79
2.50				0.31	197.00	0.80
1.50	10	Inf	0.25	0.35	172.00	0.78
3.75				0.43	97.00	0.81
1.00			0.5	0.26	316.00	0.79
2.50				0.32	179.00	0.79
1.50		8	0.25	0.34	186.00	0.79
3.75				0.43	101.00	0.79
1.00			0.5	0.25	344.00	0.80
2.50				0.32	185.00	0.80
1.50		5	0.25	0.33	198.00	0.78
3.75				0.42	104.00	0.79
1.00			0.5	0.24	362.00	0.79
2.50				0.31	192.00	0.79
1.50		3	0.25	0.31	229.00	0.79
3.75				0.41	112.00	0.78
1.00			0.5	0.23	418.00	0.79
2.50				0.30	207.00	0.79

5 Implementation in R

In this section, we demonstrate the implementation of the R function SampleSize_SMARTp for sample size calculations via a simulation study. This function is available in the R package SMARTp. The current version of this function only considers a two-stage SMART design. Figure 2 defines the SMART design.

The first three inputs of the function

```
SampleSize_SMARTp(mu, st1, dtr, regime, pow, a, rho, tau, sigma1,
              lambda, nu, sigma0, Num, p_i, c_i, a0, b0, cutoff)
```

are matrices, defined as:

- mu: mean matrix, where rows represent treatment paths and columns represent cluster sub-units (i.e., teeth) within a cluster (i.e., mouth),
- st1: stage-1 treatment matrix, where rows represent the corresponding stage-1 treatments, the 1st column includes the numbers of treatment options for responder, the 2nd column includes the numbers of treatment options for non-responders, the 3rd column are the response rates, and the 4th column includes the row numbers of matrix 'st1',
- dtr: matrix of dimension (# of DTRs X 4), the 1st column represents the DTR numbers, the 2nd column represents the corresponding treatment path numbers of responders for the corresponding DTRs in the 1st column, the 3rd column represents the corresponding treatment path numbers of the non-responders for the corresponding DTRs in the 1st column, while the 4th column represents the corresponding initial treatment.

The regime can be a single regime number if the hypothesis test is to detect the effect of that regime (e.g., Hypothesis 1), a vector of two regime numbers if the hypothesis test is to compare regimes (e.g., Hypothesis 2 or 3), or a vector of three or more regime numbers if the hypothesis test is to detect if the first regime is better than other regimes in the vector (e.g., Hypothesis 4). The power and Type-I error rates are given by pow and a respectively, with the corresponding defaults set at 0.8 and 0.05. The parameters τ and ρ, which quantify the variation and association in the CAR specification of the random effect Q_{it} are given by tau and rho, respectively, with defaults set at tau = 0.85 and rho = 0.975. The inputs sigma1, lambda and nu define the scale (σ_1), skewness (λ) and degrees of freedom (ν) parameters of the residual ϵ_{it1}, which default to sigma1 = 0.95, lambda = 0 and nu = Inf. The standard deviation σ_0 for the residual ϵ_{it0} is given by sigma0, whose default is sigma0 = 1. The rest of the parameters a_0, b_0 and c_0 from (2) are specified by a0, b0 and cutoff respectively, and their defaults are a0 = -1, b0 = 0.5 and cutoff = 0. The user can either provide the choice of a0 and b0, or the choice p_i and c_i, which are the expected proportion p_i of available teeth for patient i, and the average Pearson's correlation coefficient c_i between Y_{it} and M_{it0}, averaged over the 28 teeth for patient i, respectively. Monte Carlo estimates of the mean and variance of $\bar{Y}_i$ for each treatment path were obtained using Num random samples.

The possible outputs are summarized below:

- N, the estimated sample size;
- Del, the effect size;
- Del_std, the standardized effect size;
- ybar, the estimated regime means correspond to regime;
- Sigma, the CAR covariance matrix of Q_it;
- sig.dd, N*the variance or covariance matrix of the estimated regime means correspond to regime;
- sig.e.sq, N*the variance or covariance matrix of the difference between first and rest of estimated regime means corresponding to regime, sig.e.sq=sig.dd if the element number of regime is one;
- p_st1, the randomization probability at stage 1 for each treatment path;
- p_st2, the randomization probability at stage 2 for each treatment path;

- `res`, a vector with binary indicators represent responding or non-responding that corresponds to a treatment path;
- `ga`, the response rates of initial treatments correspond to each treatment path;
- `initr`, one column matrix with dimension of number of treatment path, the elements are the corresponding row number of `st1`.

In the following, we present the `R` code for the sample size calculation corresponding to the second row of Table 3 in Section 4.

```
#The packages required
library("mvtnorm")
library("sn")

# The SMART Design
mu=matrix(0,10,28)
mu[2,]=rep(0.5,28)
mu[4,]=rep(2,28)
mu[7,]=rep(5,28)
st1=cbind( c(1,1), c(4,4), c(0.25, 0.5), 1:2 )
dtr=cbind(1:8, c(rep(1,4),rep(6,4)), c(2,3,4,5,7,8,9,10),
          c(rep(1,4), rep(2,4)) )

## Hypothesis Test 3, with power and Type-1 error rates to be 80%
## and 20% respectively
regime=c(1,5)
pow = 0.8
a = 0.05

## Parameter values
cutoff=0; sigma1=0.95; sigma0=1; lambda=0; nu=Inf; b0=0.5; a0=-1.0;
rho=0.975;tau=0.85;

## Iteration size
Num = 1000000
```

Then, the `R` code to compute N_3, $\delta_{d_1-d_5}$, VAR$(\bar{Y}^{d_1}-\bar{Y}^{d_5})$, $\delta^{\star}_{d_1-d_5}$, VAR$(\bar{Y}^{d_1})$, VAR$(\bar{Y}^{d_5})$ and COV$(\bar{Y}^{d_1},\bar{Y}^{d_5})$ are respectively:


```
SampleSize=SampleSize_SMARTp(mu=mu, st1=st1, dtr=dtr, regime=regime,
pow=pow, a=a, rho=rho, tau=tau, sigma1=sigma1, lambda=lambda, nu=nu,
sigma0=sigma0, Num=Num, a0=a0, b0=b0, cutoff=cutoff);

N=ceiling(SampleSize$N); print(N);

Del=SampleSize$Del; print(Del);

v_d1_d5=SampleSize$sig.e.sq[1,1]/SampleSize$N;v_d1_d5;

Del_std=SampleSize$Del_std; print(Del_std);

v_d1=SampleSize$sig.dd[1,1]/SampleSize$N; print(v_d1)

v_d5=SampleSize$sig.dd[2,2]/SampleSize$N; print(v_d5)

cov_d1d5=SampleSize$sig.dd[1,2]/SampleSize$N; print(cov_d1d5)
```

6 Discussion

This paper proposes a two-stage SMART design and associated analysis plan to study a number of DTRs for managing CP. A statistical analysis plan under this design includes hypothesis testing of detecting an effect size for either a single regime, or the difference between two regimes with or without sharing an initial treatment, or even testing whether one regime is best among the embedded regimes. This paper also develops a novel sample size calculation method, accommodating typical statistical challenges observed in CP data, such as non-Gaussianity, spatial association, and non-random missingness. To the best of our knowledge, this is the *first* SMART proposal in CP research within the umbrella of *precision oral health* – a major goal in the NIH/NIDCR's Strategic Plan 2014-2019, and advances previous SMART proposals (Ghosh *et al.*, 2016; NeCamp *et al.*, 2017) considered for clustered data. Very recently, Artman *et al.* (2018) proposed rigorous sample size and power calculations factoring in the complex correlation structure between the DTRs that remain embedded within the SMART by design. However, their setup is different than what we are considering here. Our methods are fully generalizable to SMART design scenarios in other biomedical domains, where the recorded clinical endpoints exhibit similar statistical challenges as in PD studies. Furthermore, availability of open-source `R` codes from the `SMARTp` package will facilitate adaptation and modification to those scenarios.

Note, there are no real data to support the input information in the proposed sample size formulas, which some readers may view as a limitation of our methods. With precision oral health being a recently emerging field (Garaicoa-Pazmino *et al.*, 2015), this is not really a limitation of our method *per se*; it simply reflects the nascent state of the literature in this specific field. As such, we recommend considering reasonable assumptions, such as medium effect size, conservative sample sizes, etc., to elicit the input values for implementing our proposed SMART design. Additionally, these input parameters can be referenced by estimates from existing single-stage clinical trials. At any rate, our experimental design, statistical analysis plan or sample size calculation can be updated, or improved through data collection.

The proposed methodology is mainly limited to SMARTs with only two stages, and without any consideration of baseline or other covariates. In principle, our methodology can be extended to include more therapies (such as, various kinds of laser treatments), and the number of treatment stages (which leads to close monitoring of CAL changes), with each stage having more treatment types. However, such extensions may be operationally messy and/or computationally burdensome. Also, by using the Q-function

approach that minimizes squared error (NeCamp *et al.*, 2017), or maximizes likelihood (Van Der Laan and Rubin, 2006), the effect size of DTRs can be adjusted by adding baseline characteristics, such as age, gender, education, oral hygiene, etc., into the regression models in (1) and/or (2). Furthermore, the efficiency of the proposed estimator can be improved by using the corresponding doubly robust estimators, e.g., those considered by Ertefaie *et al.* (2015). Following Ertefaie *et al.* (2015) and Artman *et al.* (2018), the proposed sample size method can also be extended to detect the optimal treatment regime. These are important avenues for future research, and will be considered elsewhere.

Acknowledgements This work was supported by the grant MOE 2015-T2-2-056 from the Singapore Ministry of Education, Dr Chakraborty's start-up grant from the Duke-NUS Medical School, and the grant R01-DE024984 from the United States National Institutes of Health. The authors thank the Editor, the Associate Editor, and two reviewers, whose constructive comments led to an improved version of the manuscript. They are also thankful to researchers at the HealthPartners Institute located at Minneapolis, Minnesota for providing the motivation and the context behind this work, and also to Brian Reich from the North Carolina State University for interesting discussions.

Conflict of Interest

The authors have declared no conflict of interest. (or please state any conflicts of interest)

Appendix

A Skew-normal distribution

The statistical properties and the application of skew-normal distribution are described in Azzalini and Dalla Valle (1996) and Azzalini and Capitanio (1999) respectively. Aparecida Guedes *et al.* (2014) presents an example of applying a regression model with skew-normal errors. Here, we present a brief introduction.

Define $Z_0 \sim N(0,1)$, independent of a m-dimensional random variable $\boldsymbol{Z} = (Z_1, \ldots, Z_m)^\top$ with standardized normal marginals, and correlation matrix $\boldsymbol{\Psi}$. Suppose $\kappa_1, \ldots, \kappa_m \in (-1, 1)$, define

$$X_j = \kappa_j \mid Z_0 \mid + (1 - \kappa_j^2)^{1/2} Z_j, \tag{A-1}$$

for $j = 1, \ldots, m$, where $\kappa_j = \lambda_j/(1 + \lambda_j^2)^{1/2}$, such that $X_j \sim SN(\lambda_j)$, where 'SN' stands for skew-normal and $\boldsymbol{\lambda} \in (-\infty, \infty)$ controls skewness. The probability density function of X_j is $f(x_j; \lambda_j) = 2\phi(x_j)\Phi(\lambda_j x_j)$, for $-\infty < x_j < \infty$, where $\phi(\cdot)$ and $\Phi(\cdot)$ denote the density and cumulative distribution function (cdf) of $N(0,1)$, respectively. The joint density function of $X_1, \ldots, X_m$ is given by

$$f(\boldsymbol{x}; \boldsymbol{\theta}_x, \boldsymbol{\Omega}_x) = 2\phi_m(\boldsymbol{x}; \boldsymbol{\Omega}_x)\Phi(\boldsymbol{\theta}_x^\top \boldsymbol{x}), \tag{A-2}$$

where $\boldsymbol{x} = (x_1, \ldots, x_m)^\top$, $\phi_m(\boldsymbol{x}; \boldsymbol{\Omega}_x)$ denotes the density function of the m-dimension multivariate normal distribution with standardized marginals and correlation matrix $\boldsymbol{\Omega}_x$. We have $\boldsymbol{\theta}_x^\top = \dfrac{\boldsymbol{\lambda}^\top \boldsymbol{\Psi}^{-1} \boldsymbol{K}^{-1}}{(1 + \boldsymbol{\lambda}^\top \boldsymbol{\Psi}^{-1} \boldsymbol{\lambda})^{1/2}}$, $K = \text{diag}((1 - \kappa_1^2)^{1/2}, \ldots, (1 - \kappa_m^2)^{1/2})$, and $\boldsymbol{\Omega}_x = \boldsymbol{K}(\boldsymbol{\Psi} + \boldsymbol{\lambda}\boldsymbol{\lambda}^\top)\boldsymbol{K}$.

Define $\boldsymbol{Y} = \boldsymbol{\xi} + \boldsymbol{\omega}\boldsymbol{X}$, with $\boldsymbol{Y} = (Y_1, \ldots Y_m)^\top$, $\boldsymbol{\xi} = (\xi_1, \ldots \xi_m)^\top$ and $\boldsymbol{\omega} = \text{diag}(\omega_1, \ldots \omega_m)^\top$, where the components of $\boldsymbol{\omega}$ are assumed to be positive. The density function of $\boldsymbol{Y}$ is

$$f(\boldsymbol{y}; \boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\theta}) = 2\phi_m(\boldsymbol{y} - \boldsymbol{\xi}; \boldsymbol{\Omega})\Phi(\boldsymbol{\theta}^\top(\boldsymbol{y} - \boldsymbol{\xi})), \tag{A-3}$$

where $\boldsymbol{\Omega} = \boldsymbol{\omega}\boldsymbol{\Omega}_x\boldsymbol{\omega}$ and $\boldsymbol{\theta}^\top = \boldsymbol{\theta}_x^\top \boldsymbol{\omega}^{-1}$. Thus, $\boldsymbol{Y}$ is the m-dimensional random variable from the SN distribution, with location $\boldsymbol{\xi}$, scale $\boldsymbol{\omega}$ and skewness $\boldsymbol{\theta}$, i.e. $\boldsymbol{Y} \sim SN_m(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\theta})$. From (A-3), we have $E(\boldsymbol{Y}) = \boldsymbol{\xi} + \boldsymbol{\omega}\left(\frac{2}{\pi}\right)^{1/2}\boldsymbol{\kappa}$, $VAR(\boldsymbol{Y}) = \boldsymbol{\Omega} - \boldsymbol{\omega}^2\frac{2}{\pi}\boldsymbol{\kappa}\boldsymbol{\kappa}^\top$, and the skewness vector $SKEW(\boldsymbol{Y}) = \frac{4-\pi}{2}\frac{(\boldsymbol{\kappa}\sqrt{2/\pi})^3}{(1-2\boldsymbol{\kappa}^2/\pi)^{3/2}} = \gamma_1$

B Skew-t distribution

Skew-t random variables generated from both the skew-normal and Chi-squared variables are described in Azzalini and Capitanio (2003b). Here, $\boldsymbol{W} \sim ST_m(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\theta}, \nu)$, such that $\boldsymbol{W} = \boldsymbol{\xi} + \boldsymbol{\omega X}/\sqrt{V}$ and $\omega \boldsymbol{X} \sim SN_m(\boldsymbol{0}, \boldsymbol{\Omega}, \boldsymbol{\theta})$, where $V \sim \chi^2_\nu/\nu$ is independent of $\boldsymbol{X}$.

The density function of $\boldsymbol{W}$ is

$$f_{\boldsymbol{W}}(\boldsymbol{w}; \boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\theta}, \nu) = 2t_m(\boldsymbol{w}; \boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\theta}, \nu) T_1\left(\boldsymbol{\theta}^\top (\boldsymbol{w} - \boldsymbol{\xi}) \sqrt{\frac{\nu + m}{Q_w + \nu}}; \nu + m\right), \tag{B-1}$$

where $Q_w = (\boldsymbol{W} - \boldsymbol{\xi})^\top \boldsymbol{\Omega}^{-1} (\boldsymbol{W} - \boldsymbol{\xi})$, $t_m(\cdot; \boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\theta}, \nu)$ denotes the density function of a m-dimensional t variate with location $\boldsymbol{\xi}$, shape matrix $\boldsymbol{\Omega}$ and degrees of freedom ν, while $T_1(\cdot; \nu + m)$ denotes the cdf of an univariate student's t with degrees of freedom $\nu + m$. We can use the expression $\boldsymbol{U} = \boldsymbol{\omega X}/\sqrt{V}$ to compute the moments of $\boldsymbol{U}$, i.e. the n-th moment of $\boldsymbol{W}$ when $\boldsymbol{\xi} = 0$, is

$$E(\boldsymbol{U}^n) = \boldsymbol{\omega}^n E(\boldsymbol{X}^n) E(V^{-n/2}), \tag{B-2}$$

where

$$E(V^{-n/2}) = \frac{(\nu/2)^{n/2} \Gamma(\frac{1}{2}(\nu - n))}{\Gamma(\frac{1}{2}\nu)}$$

with $E(\boldsymbol{X}^n)$ given in Azzalini and Capitanio (1999). Thus, the mean and variance of $\boldsymbol{W}$, are, respectively,

$$E(\boldsymbol{W}) = \boldsymbol{\xi} + \boldsymbol{\omega\kappa}(\nu/\pi)^{1/2} \frac{\Gamma(\frac{1}{2}(\nu - 1))}{\Gamma(\frac{1}{2}\nu)}, \ \nu > 1,$$

$$var(\boldsymbol{W}) = \frac{\nu}{\nu - 2}\boldsymbol{\Omega} - \frac{\nu}{\pi}\left(\frac{\Gamma(\frac{1}{2}(\nu - 1))}{\Gamma(\frac{1}{2}\nu)}\right)^2 \boldsymbol{\omega}^2 \boldsymbol{\kappa\kappa}^\top, \ \nu > 2,$$

Similarly, the skewness (γ_1) and kurtosis (γ_2) for the univariate cases are

$$\gamma_1 = \mu\left[\frac{\nu(3 - \kappa^2)}{\nu - 3} - \frac{3\nu}{\nu - 2} + 2\mu^2\right]\left[\frac{\nu}{\nu - 2} - \mu^2\right]^{-3/2}, \ \nu > 3,$$

$$\gamma_2 = \left[\frac{3\nu^2}{(\nu - 2)(\nu - 4)} - \frac{4\mu^2\nu(3 - \kappa^2)}{\nu - 3} + \frac{6\mu^2\nu}{\nu - 2} - 3\mu^4\right]\left[\frac{\nu}{\nu - 2} - \mu^2\right]^{-2} - 3, \ \nu > 4,$$

where $\mu = \kappa\sqrt{\frac{\nu}{\pi}} \frac{\Gamma(\frac{1}{2}(\nu - 1))}{\Gamma(\frac{1}{2}\nu)}$.

C Sample size formula derivation

The covariance between Y_{it} and M_{it0} is

$$\begin{aligned}
&\text{COV}(\mu_i + Q_{it} + \epsilon_{it1}, a_0 + b_0 Q_{it} + \epsilon_{it0}) \\
=&\text{COV}(Q_{it}, b_0 Q_{it}) \\
=&b_0 \text{VAR}(Q_{it}) \\
=&b_0 \Sigma_{tt},
\end{aligned}$$

where Σ_{tt} is the t^{th} diagonal elements of the covariance matrix $\mathbf{\Sigma}_{28\times 28}$. The Pearson correlation coefficient is

$$\begin{aligned} c_{it} &= \frac{b_0 \text{VAR}(Q_{it})}{\sqrt{\text{VAR}(\mu_i + Q_{it} + \epsilon_{it1})\text{VAR}(a_0 + b_0 Q_{it} + \epsilon_{it0})}} \\ &= \frac{b_0 \text{VAR}(Q_{it})}{\sqrt{(\text{VAR}(Q_{it}) + \text{VAR}(\epsilon_{it1}))(b_0^2 \text{VAR}(Q_{it}) + \text{VAR}(\epsilon_{it0}))}}, \end{aligned}$$

where $\text{VAR}(Q_{it}) = \Sigma_{tt}$, $\text{VAR}(\epsilon_{it0}) = \sigma_0^2$ and $\text{VAR}(\epsilon_{it1}) = \frac{\sigma_1^2 \nu}{\nu-2} - \frac{\nu}{\pi}\left[\frac{\Gamma(0.5(\nu-1))}{\Gamma(0.5\nu)}\right]^2 \frac{\sigma_1^2\lambda^2}{1+\lambda^2}$ if $\nu < \infty$, otherwise $\text{VAR}(\epsilon_{it1}) = \sigma_1^2 - \frac{2}{\pi}\frac{\sigma_1^2\lambda^2}{1+\lambda^2}$. Now, we derive an expression for p_i, i.e.,

$$\begin{aligned} p_i =& 1 - E\left(\frac{\sum_{t=1}^{28} M_{it}}{28}\right) \\ =& 1 - \frac{1}{28}\sum_{t=1}^{28} E\left[E(M_{it} \mid Q_{it}, \epsilon_{it0})\right] \\ =& 1 - \frac{1}{28}\sum_{t=1}^{28} E\left[E(I(a_0 + b_0 Q_{it} + \epsilon_{it0} > c_0) \mid Q_{it}, \epsilon_{it0})\right] \\ =& 1 - \frac{1}{28}\sum_{t=1}^{28} E\left[I(a_0 + b_0 Q_{it} + \epsilon_{it0} > c_0)\right] \\ =& 1 - \frac{1}{28}\sum_{t=1}^{28} \text{Pr}(a_0 + b_0 Q_{it} + \epsilon_{it0} > c_0) \\ =& 1 - \frac{1}{28}\sum_{t=1}^{28} \text{Pr}\left(z > \frac{c_0 - E(a_0 + b_0 Q_{it} + \epsilon_{it0})}{\sqrt{\text{VAR}(a_0 + b_0 Q_{it} + \epsilon_{it0})}}\right) \\ =& 1 - \frac{1}{28}\sum_{t=1}^{28}\left[1 - \Phi\left(\frac{c_0 - a_0}{\sqrt{b_0^2\Sigma_{tt} + \sigma_0^2}}\right)\right] \end{aligned}$$

where $\Phi(\cdot)$ is the cdf of $z \sim$N(0,1). Now,

$$\begin{aligned} E(\bar{Y}^{d_1}) =& E\left(\frac{I(A_{i1} = a_{1i}^{d_1}, A_{i2} = [a_{2i}^{d_1 R}]^{R_i}[a_{2i}^{d_1 NR}]^{1-R_i})}{\pi_{1i}^{d_1}[\pi_{2i}^{d_1 R}]^{R_i}[\pi_{2i}^{d_1 NR}]^{1-R_i}}\bar{Y}_i\right) \\ =& E\left[E\left(\frac{I(A_{i1} = a_{1i}^{d_1}, A_{i2} = [a_{2i}^{d_1 R}]^{R_i}[a_{2i}^{d_1 NR}]^{1-R_i})}{\pi_{1i}^{d_1}[\pi_{2i}^{d_1 R}]^{R_i}[\pi_{2i}^{d_1 NR}]^{1-R_i}}\bar{Y}_i \mid R_i\right)\right] \\ =& \gamma^{d_1}\mu_{d_1 R} + (1-\gamma^{d_1})\mu_{d_1 NR}. \end{aligned}$$

In a similar way, we have $E(\bar{Y}^{d_3}) = \gamma^{d_3}\mu_{d_3 R} + (1-\gamma^{d_3})\mu_{d_3 NR}$.

According to variance decomposition, the right side of (8) is the sum of two components, which are

$$E[V(\frac{I(A_{i1} = a_{1i}^{d_1}, A_{i2} = [a_{2i}^{d_1 R}]^{R_i}[a_{2i}^{d_1 NR}]^{1-R_i})}{\pi_{1i}^{d_1}[\pi_{2i}^{d_1 R}]^{R_i}[\pi_{2i}^{d_1 NR}]^{1-R_i}}\bar{Y}_i \mid R_i)]$$

and

$$V[E(\frac{I(A_{i1}=a_{1i}^{d_1},A_{i2}=[a_{2i}^{d_1R}]^{R_i}[a_{2i}^{d_1NR}]^{1-R_i})}{\pi_{1i}^{d_1}[\pi_{2i}^{d_1R}]^{R_i}[\pi_{2i}^{d_1NR}]^{1-R_i}}\bar{Y}_i \mid R_i)].$$

The first component is

$$\begin{aligned}&E[V(\frac{I(A_{i1}=a_{1i}^{d_1},A_{i2}=[a_{2i}^{d_1R}]^{R_i}[a_{2i}^{d_1NR}]^{1-R_i})}{\pi_{1i}^{d_1}[\pi_{2i}^{d_1R}]^{R_i}[\pi_{2i}^{d_1NR}]^{1-R_i}}\bar{Y}_i \mid R_i)]\\ =&\sum_{R_i=0}^{1}V(\frac{I(A_{i1}=a_{1i}^{d_1},A_{i2}=[a_{2i}^{d_1R}]^{R_i}[a_{2i}^{d_1NR}]^{1-R_i})}{\pi_{1i}^{d_1}[\pi_{2i}^{d_1R}]^{R_i}[\pi_{2i}^{d_1NR}]^{1-R_i}}\bar{Y}_i \mid R_i)Pr(R_i),\end{aligned}$$

while the second component is

$$\begin{aligned}&V[E(\frac{I(A_{i1}=a_{1i}^{d_1},A_{i2}=[a_{2i}^{d_1R}]^{R_i}[a_{2i}^{d_1NR}]^{1-R_i})}{\pi_{1i}^{d_1}[\pi_{2i}^{d_1R}]^{R_i}[\pi_{2i}^{d_1NR}]^{1-R_i}}\bar{Y}_i \mid R_i)]\\ =&\sum_{R_i=0}^{1}E^2(\frac{I(A_{i1}=a_{1i}^{d_1},A_{i2}=[a_{2i}^{d_1R}]^{R_i}[a_{2i}^{d_1NR}]^{1-R_i})}{\pi_{1i}^{d_1}[\pi_{2i}^{d_1R}]^{R_i}[\pi_{2i}^{d_1NR}]^{1-R_i}}\bar{Y}_i \mid R_i)Pr(R_i)\\ &-(\sum_{R_i=0}^{1}E(\frac{I(A_{i1}=a_{1i}^{d_1},A_{i2}=[a_{2i}^{d_1R}]^{R_i}[a_{2i}^{d_1NR}]^{1-R_i})}{\pi_{1i}^{d_1}[\pi_{2i}^{d_1R}]^{R_i}[\pi_{2i}^{d_1NR}]^{1-R_i}}\bar{Y}_i \mid R_i)Pr(R_i))^2,\end{aligned}$$

based on the formulae $E(X)=E[E(X \mid Y)]$ and $V(X)=E(X^2)-E^2(X)$. We have $Pr(R_i=1)=\gamma^{d_1}$ or $Pr(R_i=0)=1-\gamma^{d_1}$, and

$$\begin{aligned}&V(\frac{I(A_{i1}=a_{1i}^{d_1},A_{i2}=[a_{2i}^{d_1R}])}{\pi_{1i}^{d_1}[\pi_{2i}^{d_1R}]}\bar{Y}_i \mid R_i=1)\\ =&E(\frac{I(A_{i1}=a_{1i}^{d_1},A_{i2}=[a_{2i}^{d_1R}])}{(\pi_{1i}^{d_1}[\pi_{2i}^{d_1R}])^2}\bar{Y}_i^2)-E^2(\frac{I(A_{i1}=a_{1i}^{d_1},A_{i2}=[a_{2i}^{d_1R}])}{\pi_{1i}^{d_1}[\pi_{2i}^{d_1R}]}\bar{Y}_i)\\ =&\frac{1}{\pi_{1i}^{d_1}[\pi_{2i}^{d_1R}]}E_{d_1R}(\bar{Y}_i^2)-E_{d_1R}^2(\bar{Y}_i)\\ =&\frac{1}{\pi_{1i}^{d_1}[\pi_{2i}^{d_1R}]}(\sigma_{d_1R}^2+(1-\pi_{1i}^{d_1}[\pi_{2i}^{d_1R}])\mu_{d_1R}^2),\end{aligned}$$

with μ_{d_1R}, the expectation of $\bar{Y}_i$ from d_1, with $R_i=1$, given by

$$\begin{aligned}&E_{d_1R}(\bar{Y}_i)\\ =&E_{d_1R}\left(E_{d_1R}\left(\frac{\sum_{t=1}^{28}Y_{it}(1-M_{it})}{\sum_{t=1}^{28}(1-M_{it})} \mid \boldsymbol{Q}_i,\boldsymbol{\epsilon}_{i0}\right)\right)\\ =&E_{d_1R}\left(E_{d_1R}\left(\frac{\sum_{t=1}^{28}(\mu_i+Q_{it}+\epsilon_{it1})(1-I(a_0+b_0Q_{it}+\epsilon_{it0}>0))}{\sum_{t=1}^{28}(1-I(a_0+b_0Q_{it}+\epsilon_{it0}>0))} \mid \boldsymbol{Q}_i,\boldsymbol{\epsilon}_{i0}\right)\right)\\ =&E\left(\frac{\sum_{t=1}^{28}(\mu_i\mid_{A_{i3}=1,R_i=1}+Q_{it}+E(\epsilon_{it1}))(1-I(a_0+b_0Q_{it}+\epsilon_{it0}>0))}{\sum_{t=1}^{28}(1-I(a_0+b_0Q_{it}+\epsilon_{it0}>0))}\right)\\ =&\int_{\boldsymbol{Q}_i}\int_{\boldsymbol{\epsilon}_i}\frac{\sum_{t=1}^{28}(\mu_i\mid_{A_{i3}=1,R_i=1}+Q_{it}+E(\epsilon_{it1}))(1-I(a_0+b_0Q_{it}+\epsilon_{it0}>0))}{\sum_{t=1}^{28}(1-I(a_0+b_0Q_{it}+\epsilon_{it0}>0))}f(\boldsymbol{Q}_i)f(\boldsymbol{\epsilon}_i)d\boldsymbol{\epsilon}_i d\boldsymbol{Q}_i,\end{aligned}$$

where $f(\boldsymbol{Q}_i)$ and $f(\boldsymbol{\epsilon}_i)$ are the density functions for $\boldsymbol{Q}_i$ and $\boldsymbol{\epsilon}_i$ respectively. Also, $\sigma^2_{d_1R}$, the variance of $\bar{Y}_i$ that is from d_1, with $R_i = 1$, can be written as

$$V_{d_1R}(\bar{Y}_i) = E_{d_1R}(V_{d_1R}(\bar{Y}_i \mid \boldsymbol{Q}_i, \boldsymbol{\epsilon}_{i0})) + V_{d_1R}(E_{d_1R}(\bar{Y}_i \mid \boldsymbol{Q}_i, \boldsymbol{\epsilon}_{i0})),$$

where

$$\begin{aligned}
&E_{d_1R}(V_{d_1R}(\bar{Y}_i \mid \boldsymbol{Q}_i, \boldsymbol{\epsilon}_{i0})) \\
=&E_{d_1R}\left(V_{d_1R}\left(\frac{\sum_{t=1}^{28} Y_{it}(1-M_{it})}{\sum_{t=1}^{28}(1-M_{it})} \mid \boldsymbol{Q}_i, \boldsymbol{\epsilon}_{i0}\right)\right) \\
=&E\left(V\left(\frac{\sum_{t=1}^{28}(\mu_i \mid_{A_{i3}=1,R_i=1} + Q_{it} + \epsilon_{it1})(1 - I(a_0 + b_0 Q_{it} + \epsilon_{it0} > 0))}{\sum_{t=1}^{28}(1 - I(a_0 + b_0 Q_{it} + \epsilon_{it0} > 0))}\right)\right) \\
=&E\left(\frac{\sum_{t=1}^{28}[1 - I(a_0 + b_0 Q_{it} + \epsilon_{it0} > 0)]\sigma_1^2}{\left(\sum_{t=1}^{28}[1 - I(a_0 + b_0 Q_{it} + \epsilon_{it0} > 0)]\right)^2}\right) \\
=&\int_{\boldsymbol{Q}_i}\int_{\boldsymbol{\epsilon}_i} \frac{\sum_{t=1}^{28}[1 - I(a_0 + b_0 Q_{it} + \epsilon_{it0} > 0)]\sigma_1^2}{\left(\sum_{t=1}^{28}[1 - I(a_0 + b_0 Q_{it} + \epsilon_{it0} > 0)]\right)^2} f(\boldsymbol{Q}_i) f(\boldsymbol{\epsilon}_i) d\boldsymbol{\epsilon}_i d\boldsymbol{Q}_i
\end{aligned}$$

and

$$\begin{aligned}
&V_{d_1R}(E_{d_1R}(\bar{Y}_i \mid \boldsymbol{Q}_i, \boldsymbol{\epsilon}_{i0})) \\
=&V_{d_1R}\left(E_{d_1R}\left(\frac{\sum_{t=1}^{28} Y_{it}(1-M_{it})}{\sum_{t=1}^{28}(1-M_{it})} \mid \boldsymbol{Q}_i, \boldsymbol{\epsilon}_{i0}\right)\right) \\
=&V\left(E\left(\frac{\sum_{t=1}^{28}(\mu_i \mid_{A_{i3}=1,R_i=1} + Q_{it} + \epsilon_{it1})(1 - I(a_0 + b_0 Q_{it} + \epsilon_{it0} > 0))}{\sum_{t=1}^{28}(1 - I(a_0 + b_0 Q_{it} + \epsilon_{it0} > 0))}\right)\right) \\
=&V\left(\frac{\sum_{t=1}^{28}(\mu_i \mid_{A_{i3}=1,R_i=1} + Q_{it} + E(\epsilon_{it1}))(1 - I(a_0 + b_0 Q_{it} + \epsilon_{it0} > 0))}{\sum_{t=1}^{28}(1 - I(a_0 + b_0 Q_{it} + \epsilon_{it0} > 0))}\right) \\
=&E\left[\left(\frac{\sum_{t=1}^{28}(\mu_i \mid_{A_{i3}=1,R_i=1} + Q_{it} + E(\epsilon_{it1}))(1 - I(a_0 + b_0 Q_{it} + \epsilon_{it0} > 0))}{\sum_{t=1}^{28}(1 - I(a_0 + b_0 Q_{it} + \epsilon_{it0} > 0))}\right)^2\right] \\
&- \left[E\left(\frac{\sum_{t=1}^{28}(\mu_i \mid_{A_{i3}=1,R_i=1} + Q_{it} + E(\epsilon_{it1}))(1 - I(a_0 + b_0 Q_{it} + \epsilon_{it0} > 0))}{\sum_{t=1}^{28}(1 - I(a_0 + b_0 Q_{it} + \epsilon_{it0} > 0))}\right)\right]^2 \\
=&\int_{\boldsymbol{Q}_i}\int_{\boldsymbol{\epsilon}_i}\left(\frac{\sum_{t=1}^{28}(\mu_i \mid_{A_{i3}=1,R_i=1} + Q_{it} + E(\epsilon_{it1}))(1 - I(a_0 + b_0 Q_{it} + \epsilon_{it0} > 0))}{\sum_{t=1}^{28}(1 - I(a_0 + b_0 Q_{it} + \epsilon_{it0} > 0))}\right)^2 f(\boldsymbol{Q}_i) f(\boldsymbol{\epsilon}_i) d\boldsymbol{\epsilon}_i d\boldsymbol{Q}_i \\
&- \left[\int_{\boldsymbol{Q}_i}\int_{\boldsymbol{\epsilon}_i} \frac{\sum_{t=1}^{28}(\mu_i \mid_{A_{i3}=1,R_i=1} + Q_{it} + E(\epsilon_{it1}))(1 - I(a_0 + b_0 Q_{it} + \epsilon_{it0} > 0))}{\sum_{t=1}^{28}(1 - I(a_0 + b_0 Q_{it} + \epsilon_{it0} > 0))} f(\boldsymbol{Q}_i) f(\boldsymbol{\epsilon}_i) d\boldsymbol{\epsilon}_i d\boldsymbol{Q}_i\right]^2.
\end{aligned}$$

Similarly, we have

$$V\left(\frac{I(A_{i1} = a_{1i}^{d_1}, A_{i2} = [a_{2i}^{d_1 NR}])}{\pi_{1i}^{d_1}[\pi_{2i}^{d_1 NR}]}\bar{Y}_i \mid R_i = 0\right) = \frac{1}{\pi_{1i}^{d_1}[\pi_{2i}^{d_1 NR}]}(\sigma^2_{d_1 NR} + (1 - \pi_{1i}^{d_1}[\pi_{2i}^{d_1 NR}])\mu^2_{d_1 NR})$$

, where $\mu^2_{d_1NR}$ and $\sigma^2_{d_1NR}$ are the expectation and variance of $\bar{Y}_i$ from d_1 with $R_i = 0$.

Therefore, the second component is

$$\begin{aligned}
&V[E(\frac{I(A_{i1} = a_{1i}^{d_1}, A_{i2} = [a_{2i}^{d_1R}]^{R_i}[a_{2i}^{d_1NR}]^{1-R_i})}{\pi_{1i}^{d_1}[\pi_{2i}^{d_1R}]^{R_i}[\pi_{2i}^{d_1NR}]^{1-R_i}}\bar{Y}_i \mid R_i)] \\
=&\gamma^{d_1}\mu^2_{d_1R} + (1-\gamma^{d_1})\mu^2_{d_1NR} - (\gamma^{d_1}\mu_{d_1R} + (1-\gamma^{d_1})\mu_{d_1NR})^2 \\
=&\gamma^{d_1}(1-\gamma^{d_1})(\mu_{d_1R} - \mu_{d_1NR})^2.
\end{aligned}$$

Thus, the variance formula (8) is

$$\begin{aligned}
&V(\bar{Y}^{d_1}) \\
=&\frac{1}{N}\{\frac{\gamma^{d_1}}{\pi_{1i}^{d_1}[\pi_{2i}^{d_1R}]}(\sigma^2_{d_1R} + (1-\pi_{1i}^{d_1}[\pi_{2i}^{d_1R}])\mu^2_{d_1R}) + \frac{1-\gamma^{d_1}}{\pi_{1i}^{d_1}[\pi_{2i}^{d_1NR}]}(\sigma^2_{d_1NR} + (1-\pi_{1i}^{d_1}[\pi_{2i}^{d_1NR}])\mu^2_{d_1NR}) + \\
&\gamma^{d_1}(1-\gamma^{d_1})(\mu_{d_1R} - \mu_{d_1NR})^2\}
\end{aligned}$$

while $V(\bar{Y}^{d_3})$ is

$$\begin{aligned}
&\frac{1}{N}\{\frac{\gamma^{d_3}}{\pi_{1i}^{d_3}[\pi_{2i}^{d_3R}]}(\sigma^2_{d_3R} + (1-\pi_{1i}^{d_3}[\pi_{2i}^{d_3R}])\mu^2_{d_3R}) + \frac{1-\gamma^{d_3}}{\pi_{1i}^{d_3}[\pi_{2i}^{d_3NR}]}(\sigma^2_{d_3NR} + (1-\pi_{1i}^{d_3}[\pi_{2i}^{d_3NR}])\mu^2_{d_3NR}) + \\
&\gamma^{d_3}(1-\gamma^{d_3})(\mu_{d_3R} - \mu_{d_3NR})^2\}.
\end{aligned}$$

The variance of the difference between d_1 and d_3 is

$$V(\bar{Y}^{d_1} - \bar{Y}^{d_3}) = V(\bar{Y}^{d_1}) + V(\bar{Y}^{d_3}) - 2\text{COV}(\bar{Y}^{d_1}, \bar{Y}^{d_3}). \tag{C-1}$$

The covariance between $\bar{Y}^{d_1}$ and $\bar{Y}^{d_3}$ is

$$
\begin{aligned}
&\mathrm{COV}\left(\bar{Y}^{d_1},\bar{Y}^{d_3}\right)\\
=&\frac{1}{N^2}\mathrm{COV}\left(\sum_{i=1}^{N}W_i^{d_1}\bar{Y}_i,\sum_{i=1}^{N}W_i^{d_3}\bar{Y}_i\right)\\
=&\frac{1}{N^2}\mathrm{COV}\left(\sum_{i=1}^{N}W_i^{d_1}\bar{Y}_i(R_i+(1-R_i)),\sum_{i=1}^{N}W_i^{d_3}\bar{Y}_i(R_i+(1-R_i))\right)\\
=&\frac{1}{N^2}\mathrm{COV}\left(\sum_{i=1}^{N}W_i^{d_1}\bar{Y}_iR_i+W_i^{d_1}\bar{Y}_i(1-R_i),\sum_{i=1}^{N}W_i^{d_3}\bar{Y}_iR_i+W_i^{d_3}\bar{Y}_i(1-R_i)\right)\\
=&\frac{1}{N^2}\mathrm{COV}\left(\sum_{i=1}^{N}W_i^{d_1}\bar{Y}_iR_i+\sum_{i=1}^{N}W_i^{d_1}\bar{Y}_i(1-R_i),\sum_{i=1}^{N}W_i^{d_3}\bar{Y}_iR_i+\sum_{i=1}^{N}W_i^{d_3}\bar{Y}_i(1-R_i)\right)\\
=&\frac{1}{N^2}\mathrm{COV}\left(\sum_{i=1}^{N}W_i^{d_1}\bar{Y}_iR_i,\sum_{i=1}^{N}W_i^{d_3}\bar{Y}_iR_i\right)+\frac{1}{N^2}\mathrm{COV}\left(\sum_{i=1}^{N}W_i^{d_1}\bar{Y}_iR_i,\sum_{i=1}^{N}W_i^{d_3}\bar{Y}_i(1-R_i)\right)\\
&+\frac{1}{N^2}\mathrm{COV}\left(\sum_{i=1}^{N}W_i^{d_1}\bar{Y}_i(1-R_i),\sum_{i=1}^{N}W_i^{d_3}\bar{Y}_iR_i\right)+\frac{1}{N^2}\mathrm{COV}\left(\sum_{i=1}^{N}W_i^{d_1}\bar{Y}_i(1-R_i),\sum_{i=1}^{N}W_i^{d_3}\bar{Y}_i(1-R_i)\right)\\
=&\frac{1}{N^2}\left[\sum_{i=1}^{N}\mathrm{COV}\left(W_i^{d_1}\bar{Y}_iR_i,W_i^{d_3}\bar{Y}_iR_i\right)+\sum_{i\neq j}\mathrm{COV}\left(W_i^{d_1}\bar{Y}_iR_i,W_j^{d_3}\bar{Y}_jR_j\right)\right]\\
&+\frac{1}{N^2}\left[\sum_{i=1}^{N}\mathrm{COV}\left(W_i^{d_1}\bar{Y}_iR_i,W_i^{d_3}\bar{Y}_i(1-R_i)\right)+\sum_{i\neq j}\mathrm{COV}\left(W_i^{d_1}\bar{Y}_iR_i,W_j^{d_3}\bar{Y}_j(1-R_j)\right)\right]\\
&+\frac{1}{N^2}\left[\sum_{i=1}^{N}\mathrm{COV}\left(W_i^{d_1}\bar{Y}_i(1-R_i),W_i^{d_3}\bar{Y}_iR_i\right)+\sum_{i\neq j}\mathrm{COV}\left(W_i^{d_1}\bar{Y}_i(1-R_i),W_j^{d_3}\bar{Y}_jR_j\right)\right]\\
&+\frac{1}{N^2}\left[\sum_{i=1}^{N}\mathrm{COV}\left(W_i^{d_1}\bar{Y}_i(1-R_i),W_i^{d_3}\bar{Y}_i(1-R_i)\right)+\sum_{i\neq j}\mathrm{COV}\left(W_i^{d_1}\bar{Y}_i(1-R_i),W_j^{d_3}\bar{Y}_j(1-R_j)\right)\right]\\
=&\frac{1}{N}\mathrm{COV}\left(W_i^{d_1}\bar{Y}_iR_i,W_i^{d_3}\bar{Y}_iR_i\right)+\frac{1}{N}\mathrm{COV}\left(W_i^{d_1}\bar{Y}_iR_i,W_i^{d_3}\bar{Y}_i(1-R_i)\right)\\
&+\frac{1}{N}\mathrm{COV}\left(W_i^{d_1}\bar{Y}_i(1-R_i),W_i^{d_3}\bar{Y}_iR_i\right)+\frac{1}{N}\mathrm{COV}\left(W_i^{d_1}\bar{Y}_i(1-R_i),W_i^{d_3}\bar{Y}_i(1-R_i)\right),
\end{aligned}
$$

where

$$
\begin{aligned}
&\text{COV}\left(W_i^{d_1}\bar{Y}_i R_i, W_i^{d_3}\bar{Y}_i R_i\right)\\
=&E(W_i^{d_1}\bar{Y}_i R_i W_i^{d_3}\bar{Y}_i R_i) - E(W_i^{d_1}\bar{Y}_i R_i)E(W_i^{d_3}\bar{Y}_i R_i)\\
=&E[E(W_i^{d_1}\bar{Y}_i R_i W_i^{d_3}\bar{Y}_i R_i) \mid R_i] - E[E(W_i^{d_1}\bar{Y}_i R_i) \mid R_i]E[E(W_i^{d_3}\bar{Y}_i R_i) \mid R_i]\\
=&\frac{\gamma^{d_1}}{\pi_{1i}^{d_1}\pi_{2i}^{d_1 R}}E_{d_1 R}(\bar{Y}_i^2)\\
&-\gamma^{d_1}\gamma^{d_3}E_{d_1 R}(\bar{Y}_i)E_{d_3 R}(\bar{Y}_i)\\
=&\frac{\gamma^{d_1}}{\pi_{1i}^{d_1}\pi_{2i}^{d_1 R}}\left[\sigma_{d_1 R}^2+\mu_{d_1 R}^2\right]-\gamma^{d_1}\gamma^{d_3}\mu_{d_1 R}\mu_{d_3 R},
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
&\text{COV}\left(W_i^{d_1}\bar{Y}_i R_i, W_i^{d_3}\bar{Y}_i (1-R_i)\right)\\
=&E[E(W_i^{d_1}\bar{Y}_i R_i W_i^{d_3}\bar{Y}_i (1-R_i)) \mid R_i] - E[E(W_i^{d_1}\bar{Y}_i R_i) \mid R_i]E[E(W_i^{d_3}\bar{Y}_i (1-R_i)) \mid R_i]\\
=&-\gamma^{d_1}(1-\gamma^{d_3})\mu_{d_1 R}\mu_{d_3 NR},
\end{aligned}
$$

$$
\begin{aligned}
&\text{COV}\left(W_i^{d_1}\bar{Y}_i (1-R_i), W_i^{d_3}\bar{Y}_i R_i\right)\\
=&E[E(W_i^{d_1}\bar{Y}_i (1-R_i) W_i^{d_3}\bar{Y}_i R_i) \mid R_i] - E[E(W_i^{d_1}\bar{Y}_i (1-R_i)) \mid R_i]E[E(W_i^{d_3}\bar{Y}_i R_i) \mid R_i]\\
=&-\gamma^{d_3}(1-\gamma^{d_1})\mu_{d_1 NR}\mu_{d_3 R},
\end{aligned}
$$

$$
\begin{aligned}
&\text{COV}\left(W_i^{d_1}\bar{Y}_i (1-R_i), W_i^{d_3}\bar{Y}_i (1-R_i)\right)\\
=&E[E(W_i^{d_1}\bar{Y}_i (1-R_i) W_i^{d_3}\bar{Y}_i (1-R_i)) \mid R_i] - E[E(W_i^{d_1}\bar{Y}_i (1-R_i)) \mid R_i]E[E(W_i^{d_3}\bar{Y}_i (1-R_i)) \mid R_i]\\
=&-(1-\gamma^{d_1})(1-\gamma^{d_3})\mu_{d_1 NR}\mu_{d_3 NR}.
\end{aligned}
$$

Thus, COV $\left(\bar{Y}^{d_1}, \bar{Y}^{d_3}\right)$ is

$$
\begin{aligned}
&\text{COV}\left(\bar{Y}^{d_1}, \bar{Y}^{d_3}\right)\\
=&\frac{1}{N}\{\frac{\gamma^{d_1}}{\pi_{1i}^{d_1}\pi_{2i}^{d_1 R}}(\sigma_{d_1 R}^2+\mu_{d_1 R}^2)-\gamma^{d_1}\gamma^{d_3}\mu_{d_1 R}\mu_{d_3 R}\\
&-\gamma^{d_1}(1-\gamma^{d_3})\mu_{d_1 R}\mu_{d_3 NR}\\
&-\gamma^{d_3}(1-\gamma^{d_1})\mu_{d_1 NR}\mu_{d_3 R}\\
&-(1-\gamma^{d_1})(1-\gamma^{d_3})\mu_{d_1 NR}\mu_{d_3 NR}\}.
\end{aligned}
$$

Therefore, the variance of regime means differences between d_1 and d_3 is

$$
\begin{aligned}
& V(\bar{Y}^{d_1} - \bar{Y}^{d_3}) \\
= & \frac{1}{N}\{\frac{\gamma^{d_1}}{\pi_{1i}^{d_1}[\pi_{2i}^{d_1 R}]}(\sigma^2_{d_1 R} + (1 - \pi_{1i}^{d_1}[\pi_{2i}^{d_1 R}])\mu^2_{d_1 R}) + \\
& \frac{1-\gamma^{d_1}}{\pi_{1i}^{d_1}[\pi_{2i}^{d_1 NR}]}(\sigma^2_{d_1 NR} + (1 - \pi_{1i}^{d_1}[\pi_{2i}^{d_1 NR}])\mu^2_{d_1 NR}) + \\
& \gamma^{d_1}(1-\gamma^{d_1})(\mu_{d_1 R} - \mu_{d_1 NR})^2 + \\
& \frac{\gamma^{d_3}}{\pi_{1i}^{d_3}[\pi_{2i}^{d_3 R}]}(\sigma^2_{d_3 R} + (1 - \pi_{1i}^{d_3}[\pi_{2i}^{d_3 R}])\mu^2_{d_3 R}) + \\
& \frac{1-\gamma^{d_3}}{\pi_{1i}^{d_3}[\pi_{2i}^{d_3 NR}]}(\sigma^2_{d_3 NR} + (1 - \pi_{1i}^{d_3}[\pi_{2i}^{d_3 NR}])\mu^2_{d_3 NR}) + \\
& \gamma^{d_3}(1-\gamma^{d_3})(\mu_{d_3 R} - \mu_{d_3 NR})^2 \\
& - 2[\frac{\gamma^{d_1}}{\pi_{1i}^{d_1}\pi_{2i}^{d_1 R}}(\sigma^2_{d_1 R} + \mu^2_{d_1 R}) - \gamma^{d_1}\gamma^{d_3}\mu_{d_1 R}\mu_{d_3 R} \\
& - \gamma^{d_1}(1-\gamma^{d_3})\mu_{d_1 R}\mu_{d_3 NR} \\
& - \gamma^{d_3}(1-\gamma^{d_1})\mu_{d_1 NR}\mu_{d_3 R} \\
& - (1-\gamma^{d_1})(1-\gamma^{d_3})\mu_{d_1 NR}\mu_{d_3 NR}]\}.
\end{aligned}
$$

Similarly, we can also derive the expression for COV $\left(\bar{Y}^{d_1}, \bar{Y}^{d_5}\right)$, where d_1 and d_5 does not share an initial treatment, i.e.

$$
\begin{aligned}
& \text{COV}\left(\bar{Y}^{d_1}, \bar{Y}^{d_5}\right) \\
= & \frac{1}{N}\{-\gamma^{d_1}\gamma^{d_5}\mu_{d_1 R}\mu_{d_5 R} \\
& - \gamma^{d_1}(1-\gamma^{d_5})\mu_{d_1 R}\mu_{d_5 NR} \\
& - \gamma^{d_5}(1-\gamma^{d_1})\mu_{d_1 NR}\mu_{d_5 R} \\
& - (1-\gamma^{d_1})(1-\gamma^{d_5})\mu_{d_1 NR}\mu_{d_5 NR}\}.
\end{aligned}
$$

Thus, all the variances of $\bar{Y}^{d_1} - \bar{Y}^{d_2}$,... and $\bar{Y}^{d_1} - \bar{Y}^{d_8}$ can be derived in the similar ways as $V(\bar{Y}^{d_1} - \bar{Y}^{d_3})$ or $V(\bar{Y}^{d_1} - \bar{Y}^{d_5})$. So do the covariances among $\bar{Y}^{d_1} - \bar{Y}^{d_2}$,... and $\bar{Y}^{d_1} - \bar{Y}^{d_8}$, e.g.,

$$
\begin{aligned}
& \text{COV}\left(\bar{Y}^{d_1} - \bar{Y}^{d_2}, \bar{Y}^{d_1} - \bar{Y}^{d_3}\right) \\
= & \text{COV}\left(\bar{Y}^{d_1}, \bar{Y}^{d_1}\right) - \text{COV}\left(\bar{Y}^{d_1}, \bar{Y}^{d_3}\right) - \text{COV}\left(\bar{Y}^{d_1}, \bar{Y}^{d_2}\right) + \text{COV}\left(\bar{Y}^{d_2}, \bar{Y}^{d_3}\right).
\end{aligned}
$$

Now, equation (14) can be derived as follows

$$\Pr\left(\frac{\hat{\delta}_{d_1-d_2}}{\sqrt{2\sigma_{d_1-d_2}/N_4}} > z_\alpha, \ldots, \frac{\hat{\delta}_{d_1-d_8}}{\sqrt{2\sigma_{d_1-d_8}/N_4}} > z_\alpha\right) = 1-\beta = \text{power}$$

$$\Longrightarrow \Pr\left(\frac{\hat{\delta}_{d_1-d_2}-\delta_{d_1-d_2}}{\sqrt{2\sigma_{d_1-d_2}/N_4}} > z_\alpha - \frac{\delta_{d_1-d_2}}{\sqrt{2\sigma_{d_1-d_2}/N_4}}, \ldots, \frac{\hat{\delta}_{d_1-d_8}-\delta_{d_1-d_8}}{\sqrt{2\sigma_{d_1-d_8}/N_4}} > z_\alpha - \frac{\delta_{d_1-d_8}}{\sqrt{2\sigma_{d_1-d_8}/N_4}}\right) = \text{power}$$

$$\Longrightarrow \Pr\left(\frac{\hat{\delta}_{d_1-d_2}-\delta_{d_1-d_2}}{\sqrt{2\sigma_{d_1-d_2}/N_4}} > z_\alpha - \sqrt{N_4/2}\delta^*_{d_1-d_2}, \ldots, \frac{\hat{\delta}_{d_1-d_8}-\delta_{d_1-d_8}}{\sqrt{2\sigma_{d_1-d_8}/N_4}} > z_\alpha - \sqrt{N_4/2}\delta^*_{d_1-d_8}\right) = \text{power}$$

$$\Longrightarrow \Pr\left(Z_{d_1-d_2} \le -(z_\alpha - \delta^*_{d_1-d_2}\sqrt{N_4/2}), \ldots, Z_{d_1-d_8} \le -(z_\alpha - \delta^*_{d_1-d_8}\sqrt{N_4/2})\right) = \text{power}$$

D Proof of Theorem 3.1

Proof:
The proof of consistency requires the result of strong law of large numbers, such that $\hat{\delta}_{d_1-d_3} = \bar{Y}_{d_1} - \bar{Y}_{d_3} = \frac{1}{N_2}\sum_{i=1}^{N_2} W_i^{d_1}\bar{Y}_i - \frac{1}{N_2}\sum_{i=1}^{N_2} W_i^{d_3}\bar{Y}_i \to \mu_{d_10} - \mu_{d_30}$, almost surely, and uniformly for μ_{d_1} and $\mu_{d_3} \in \Theta$ as $N_2 \to \infty$ and $\delta_{(d_1-d_3)0}$ being the unique expected value of $\hat{\delta}_{d_1-d_3}$ due to Assumption 2. To prove the asymptotic normality result, we have

$$\begin{aligned}&\sqrt{N_2}(\hat{\delta}_{d_1-d_3} - \delta_{(d_1-d_3)0})\\ =&\sqrt{N_2}\left[\frac{1}{N_2}\sum_{i=1}^{N_2} W_i^{d_1}\bar{Y}_i - \frac{1}{N_2}\sum_{i=1}^{N_2} W_i^{d_3}\bar{Y}_i - (\mu_{d_10} - \mu_{d_30})\right]\end{aligned}$$

with $E\left(\frac{1}{N_2}\sum_{i=1}^{N_2} W_i^{d_1}\bar{Y}_i\right) = \mu_{d_10}$, $E\left(\frac{1}{N_2}\sum_{i=1}^{N_2} W_i^{d_3}\bar{Y}_i\right) = \mu_{d_30}$, $E\left(\hat{\delta}_{d_1-d_3}\right) = \delta_{(d_1-d_3)0}$ and $\text{var}(\hat{\delta}_{d_1-d_3}) = 2\sigma^2_{d_1-d_3}/N_2$ due to Assumptions 1 and 2. Note that $\sigma^2_{d_1-d_3}$ can be computed by equation (10). Thus, by the central limit theorem, $\sqrt{N_2}(\hat{\delta}_{d_1-d_3} - \delta_{(d_1-d_3)0})$ converges in distribution to Normal$(0, 2\sigma^2_{d_1-d_3})$.

References

[1] Albert, P. S. (2019). Shared random parameter models: A legacy of the biostatistics program at the National Heart, Lung, and Blood Institute. *Statistics in Medicine* **38**, 501–511.

[2] Aparecida Guedes, T., Rossi, R. M., Tozzo Martins, A. B., Janeiro, V., and Pedroza Carneiro, J. W. (2014). Applying regression models with skew-normal errors to the height of bedding plants of Stevia rebaudiana (Bert) Bertoni. *Acta Scientiarum. Technology* **36**, 463–468.

[3] Artman, W., Nahum-Shani, I., Wu, T., Mckay, J., and Ertefaie, A. (2018). Power analysis in a SMART design: sample size estimation for determining the best embedded dynamic treatment regime. *Biostatistics* , 1–17.

[4] Azarpazhooh, A., Shah, P. S., Tenenbaum, H. C., and Goldberg, M. B. (2010). The effect of photodynamic therapy for periodontitis: a systematic review and meta-analysis. *Journal of Periodontology* **81**, 4–14.

[5] Azzalini, A. and Capitanio, A. (1999). Statistical applications of the multivariate skew normal distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **61**, 579–602.

[6] Azzalini, A. and Capitanio, A. (2003a). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew-t distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **65**, 367–389.

[7] Azzalini, A. and Capitanio, A. (2003b). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew-t distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **65**, 367—389.

[8] Azzalini, A. and Dalla Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika* **83**, 715–726.

[9] Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2014). *Hierarchical Modeling and Analysis for Spatial Data*. Chapman and Hall/CRC, 2 edition.

[10] Beck, J. D., Elter, J. R., Heiss, G., Couper, D., Mauriello, S. M., *et al.* (2001). Relationship of periodontal disease to carotid artery intima-media wall thickness: the atherosclerosis risk in communities (ARIC) study. *Arteriosclerosis, Thrombosis, and Vascular Biology* **21**, 1816–1822.

[11] Breslin, F. C., Sobell, M. B., Sobell, L. C., Cunningham, J. A., Sdao-Jarvie, K., *et al.* (1998). Problem drinkers: evaluation of a stepped-care approach. *Journal of Substance Abuse* **10**, 217–232.

[12] Brooner, R. and Kidorf, M. (2002). Using behavioral reinforcement to improve methadone treatment participation. *Science and Practice Perspectives* **1**, 38–47.

[13] Cao, W., Tsiatis, A., and Davidian, M. (2009). Improving efficiency and robustness of the doubly robust estimator for a population mean with incomplete data. *Biometrika* **96**, 723–734.

[14] Chakraborty, B. and Moodie, E. (2013). *Statistical Methods for Dynamic Treatment Regimes*. Springer, New York.

[15] Chakraborty, B., Murphy, S., and Strecher, V. (2010). Inference for non-regular parameters in optimal dynamic treatment regimes. *Statistical Methods in Medical Research* **19**, 317—-343.

[16] Eke, P., Page, R., Wei, L., Thornton-Evans, G., and Genco, R. (2012). Update of the Case Definitions for Population-Based Surveillance of Periodontitis. *Journal of Periodontology* **83**, 1449–1454.

[17] Ertefaie, A., Wu, T., Lynch, K., and Nahum-Shani, I. (2015). Identifying a set that contains the best dynamic treatment regimes. *Biostatistics* **17**, 135–148.

[18] Fernandes, J. K., Wiegand, R. E., Salinas, C. F., Grossi, S. G., Sanders, J. J., *et al.* (2009). Periodontal disease status in Gullah African Americans with Type-2 diabetes living in South Carolina. *Journal of Periodontology* **80**, 1062–1068.

[19] Garaicoa-Pazmino, C., Decker, A. M., and Polverini, P. J. (2015). Personalized Medicine Approaches to the Prevention, Diagnosis, and Treatment of Chronic Periodontitis. In Polverini, P. J. (ed.), *Personalized Oral Health Care*. Springer, pages 99–112.

[20] Garcia, I., Kuska, R., and Somerman, M. (2013). Expanding the Foundation for Personalized Medicine: Implications and Challenges for Dentistry. *Journal of Dental Research Clinical Research Supplement* **92**, 3S–10S.

[21] Ghosh, P., Cheung, Y., and Chakraborty, B. (2016). Sample size calculations for clustered SMART designs. In Kosorok, M. and Moodie, E. (eds.), *Adaptive Treatment Strategies in Practice: Planning Trials and Analyzing Data for Personalized Medicine*, chapter 5. ASA-SIAM Statistics and Applied Probability Series, Philadelphia, PA, pages 55–68.

[22] Glasgow, M., Engel, B., and D'Lugoff, B. (1989). A controlled study of a standardized behavioural stepped treatment for hypertension. *Psychosomatic Medicine* **51**, 10–26.

[23] Grossi, S. G., Skrepcinski, F. B., DeCaro, T., Robertson, D. C., Ho, A. W., *et al.* (1997). Treatment of periodontal disease in diabetics reduces glycated hemoglobin. *Journal of Periodontology* **68**, 713–719.

[24] Herrera, D. (2016). Scaling and root planning is recommended in the nonsurgical treatment of chronic periodontitis. *Journal of Evidence-Based Dental Practice* **16**, 56–58.

[25] John, M., Michalowicz, B., Kotsakis, G., and Chu, H. (2017). Network meta-analysis of studies included in the Clinical Practice Guideline on the nonsurgical treatment of chronic periodontitis. *Journal of Clinic Peridontology* **44**, 603–611.

[26] Lavori, P. W. and Dawson, R. (2004). Dynamic treatment regimes: Practical design considerations. *Clinical Trials* **1**, 9–20.

[27] Lei, H., Nahum-Shani, I., Lynch, K., Oslin, D., and Murphy, S. A. (2012). A "SMART" Design for Building Individualized Treatment Sequences. *Annual Review of Clinical Psychology* **8**, 21–48.

[28] Liu, C.-M., Hou, L.-T., Wong, M.-Y., and Lan, W.-H. (1999). Comparison of Nd: YAG laser versus scaling and root planing in periodontal therapy. *Journal of Periodontology* **70**, 1276–1282.

[29] Murphy, S. A. (2003). Optimal dynamic treatment regimes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **65**, 331–366.

[30] Murphy, S. A. (2005). An experimental design for the development of adaptive treatment strategies. *Statistics in Medicine* **24**, 1455–1481.

[31] Murphy, S. A., van der Laan, M. J., Robins, J. M., and Group, C. P. P. R. (2001). Marginal mean models for dynamic regimes. *Journal of the American Statistical Association* **96**, 1410–1423.

[32] Murphy, S. A. and McKay, J. R. (2004). Adaptive treatment strategies: An emerging approach for improving treatment effectiveness. *Clinical Science* **12**, 7–13.

[33] NeCamp, T., Kilbourne, A., and Almirall1, D. (2017). Comparing cluster-level dynamic treatment regimens using sequential, multiple assignment, randomized trials: Regression estimation and sample size considerations. *Statistical Methods in Medical Research* **26**, 1572–1589.

[34] Nicholls, C. (2003). Periodontal disease incidence, progression and rate of tooth loss in a general dental practice: The results of a 12-year retrospective analysis of patient's clinical records. *British Dental Journal* **194**, 485–488.

[35] Porteous, M. S. and Rowe, D. J. (2014). Adjunctive Use of the Diode Laser in Non-surgical Periodontal Therapy: Exploring the Controversy. *The Journal of Dental Hygiene* **88**, 78–86.

[36] Reich, B. J. and Bandyopadhyay, D. (2010). A latent factor model for spatial data with informative missingness. *The Annals of Applied Statistics* **4**, 439–459.

[37] Reich, B., Bandyopadhyay, D., and Bondell, H. (2013). A Nonparametric Spatial Model for Periodontal Data with Nonrandom Missingness. *Journal of the American Statistical Association* **108**, 820–831.

[38] Robins, J. M. (2004). Optimal Structural Nested Models for Optimal Sequential Decisions. In Lin, D. and Heagerty, P. (eds.), *Proceedings of the Second Seattle Symposium in Biostatistics: Analysis of Correlated Data*, volume 179 of *Lecture Notes in Statistics*, chapter 11. Springer, New York, pages 189–326.

[39] Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1994a). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association* **89**, 846–866.

[40] Robins, J., Rotnitzky, A., and Zhao, L. (1994b). Estimation of Regression Coefficients When Some Regressors are not Always Observed. *Journal of the American Statistical Association* **89**, 846–866.

[41] Sgolastra, F., Petrucci, A., Gatto, R., and Monaco, A. (2012). Efficacy of Er:YAG laser in the treatment of chronic periodontitis: systematic review and meta-analysis. *Lasers in Medical Science* **27**, 661–673.

[42] Smiley, C., Tracy, S., Abt, E., Michalowicz, B., John, M., *et al.* (2015). Systematic review and meta-analysis on the nonsurgical treatment of chronic periodontitis by means of scaling and root planing with or without adjuncts. *Journal of American Dental Association* **146**, 508–524.

[43] Thornton-Evans, G., Eke, P., Wei, L., Palmer, A., Moeti, R., *et al.* (2013). Periodontitis among adults aged $\geq$30 years - United States, 2009-2010. *Centers for Disease Control and Prevention. Morbidity and mortality weekly report* **62**, 129–135.

[44] Untzer, J., Katon, W., Williams, J., Callahan, C., Harpole, L., *et al.* (2001). Improving primary care for depression in late life: the design of a multicenter randomized trial. *Medical Care* **39**, 785—-799.

[45] Van Der Laan, M. and Rubin, D. (2006). Targeted Maximum Likelihood Learning. Working Paper Series Working Paper 213, U.C. Berkeley Division of Biostatistics.

[46] Vonesh, E. F., Greene, T., and Schluchter, M. D. (2006). Shared parameter models for the joint analysis of longitudinal data and event times. *Statistics in Medicine* **25**, 143–163.

[47] Wang, Z., Zhou, X., Zhang, J., Zhang, L., Song, Y., *et al.* (2009). Periodontal health, oral health behaviours, and chronic obstructive pulmonary disease. *Journal of Clinical Periodontology* **36**, 750–755.

[48] Wiebe, C. B. and Putnins, E. E. (2000). The Periodontal Disease Classification System of the American Academy of Periodontology – An Update. *Journal of the Canadian Dental Association* **66**, 594–597.

[49] Workgroup, A. (2011). American Academy of Periodontology Statement on the efficacy of lasers in the non-surgical treatment of inflammatory periodontal disease. *Journal of Periodontology* **82**, 513–514.

[50] Zhao, Y., Yin, Y., Tao, L., Nie, P., Tang, Y., *et al.* (2014). Er:YAG laser versus scaling and root planing as alternative or adjuvant for chronic periodontitis treatment: a systematic review. *Journal of Clinical Periodontology* **41**, 1069–1079.

