Sequential Multiple Assignment Randomized Trial (SMART) To Identify Optimal Sequences Of Telemedicine Interventions For Improving Initiation of Insulin Therapy: A Simulation Study

Xiaoxi Yan (✉ xiaoxi.yan@u.duke-nus.edu)
Duke-NUS Medical School

David B. Matchar
Duke-NUS Medical School

Aastha Goel
Duke-NUS Medical School

Nirmali Sivapragasam
Duke-NUS Medical School

John P. Ansah
Duke-NUS Medical School

Bibhas Chakraborty
Duke-NUS Medical School

Abstract

Background: To examine the value of a Sequential Multiple Assignment Randomized Trial (SMART) design compared to a conventional randomized control trial (RCT) for telemedicine strategies to support titration of insulin therapy for Type 2 Diabetes Mellitus (T2DM) patients new to insulin.

Methods: Microsimulation models were created in R using a synthetic sample based on primary data from 63 subjects enrolled in a pilot study of a smartphone application (App), *Diabetes Pal* compared to a nurse-based telemedicine strategy (Nurse). For comparability, the SMART and an RCT design were constructed to allow comparison of four (embedded) adaptive interventions (AIs).

Results: In the base case scenario, the SMART has similar overall mean expected HbA1c and cost per subject compared with RCT, for sample size of n = 100 over 10000 simulations. SMART has lower (better) standard deviations of the mean expected HbA1c per AI, and higher efficiency of choosing the correct AI across various sample sizes. The differences between SMART and RCT become apparent as sample size decreases. For both trial designs, the threshold value at which a subject was deemed to have been responsive at an intermediate point in the trial had an optimal choice (i.e., the sensitivity curve had a U-shape). SMART design dominates the RCT, in the overall mean HbA1c (lower value) when the threshold value is close to optimal.

Conclusions: SMART is suited to evaluating the efficacy of different sequences of treatment options, in addition to the advantage of providing information on optimal treatment sequences.

Introduction

A major objective of clinical trials, particularly randomized controlled trials (RCTs) is to identify which of two or more therapies is most effective. However, people often differ in their response to the same intervention. When a treatment that works for most people based on an RCT is not effective for a particular patient, in clinical practice the next step typically is to try something else. The next choice in this "trial and error" process would, ideally, be informed by evidence. However, clinical trials in which individuals are randomized to sequences of treatment strategies are seldom used [1].

An alternative to an idiosyncratic series of choices are decision rules such as those embodied in guidelines developed by medical professional organizations: a combination of expert opinion, behavioral, psychosocial and biological theories, and observational studies to formulate adaptive treatment algorithms, or adaptive interventions (AIs) [2, 3]. While clinical guidelines may reduce variability from practice to practice, they do not alleviate the scientific uncertainty about which sequence is actually optimal. The recommendations become the subject of potential future research.

Experimental trial designs have been proposed for development and optimization of treatment sequences. One such design is the Sequential Multiple Assignment Randomized Trial (SMART) [2, 4]. Adaptive interventions are treatment algorithms wherein treatment is sequentially modified over time

based on individual's response. The rationale is that by adjusting the treatment type and level as a function of time-dependent measures such as response to the past treatment, the long-term outcome is optimized [2, 5].

Most experience with SMARTs has been limited to mental health and behavioral sciences [2, 4], and Phase 2 trials in oncology [6]. SMART is particularly attractive in cancer therapy as sequential treatment based on intermediate response is already well-established. However, SMART has potential value to scientifically address problems in a wide range of contexts, including the use of technology such as telemedicine to encourage health-promoting behaviors [7].

Telemedicine is the provision of healthcare services and the exchange of healthcare information using information and communication technology across distances [8, 9]. What is becoming evident in this field is that "one size does not fit all". Studies have shown that telemedicine interventions are more likely to have a positive effect on users' self-efficacy, knowledge relevant to their condition, and behavioral and clinical outcomes [10]. However, not all patients are receptive to a particular mode of delivery. A key to establishing the effective and cost-effective application of telemedicine is understanding how these approaches fit into real-world care, in particular as part of a sequence that maximizes the proportion of patients who ultimately respond to good effect.

With this in mind, we sought to examine the value of a SMART design compared to an RCT for two telemedicine strategies to support titration of insulin therapy for Type 2 Diabetes Mellitus (T2DM) patients new to insulin: (1) a largely self-contained smartphone app, Diabetes Pal [11] and (2) a nurse-based telephone consultation service, SingHealth Polyclinics' (SHP) Insulin Initiation Telecare Program (see the Methods section for details about these two telemedicine modalities). For comparability, the SMART and an RCT designs were constructed to allow comparison of various sequences of the two telemedicine strategies. The basis for this comparison is microsimulation using data derived from a pilot clinical trial of Diabetes Pal [11]. We sought to demonstrate the impact of the two trial designs on improvement in chronic blood glucose control as measured by change in glycated hemoglobin (HbA1c), and trial cost for the study population. In sensitivity analysis we examined how these measures of value were affected by various aspects of trial design, including the operating characteristics of the measure of responsiveness to initial treatment measure used to determine whether to continue or switch treatment.

Methods

Overview of the Simulation Study

The purpose of our simulation study was to conduct a head-to-head comparison between two design approaches intended to identify the optimal sequence of the two telemedicine modalities for titration of insulin dose in insulin-naïve diabetic patients. Although it is impractical to compare the design approaches directly using the same set of participants in real-life empirical studies, such comparisons are possible in a computer simulation. Specifically, for this study we developed a microsimulation created in

R 3.6.1 [12]. The synthetic subjects were generated based on the characteristics of the real subjects in the pilot study of the *Diabetes Pal* app [11].

Two Telemedicine Intervention Modalities

Here we briefly describe the telemedicine intervention modalities that were compared in the pilot study, and informed by that, were considered in our simulation study.

(a) *SingHealth Polyclinics' Insulin Initiation Telecare Program (Nurse)*: The program was designed to support insulin initiation for patients with T2DM at the primary care practices of SingHealth, the largest public healthcare group in Singapore. Designated primary care nurses were trained as care managers, to assist patients with insulin initiation via weekly telephone consultations. These consultations included checking of current insulin dose and presence of symptoms of hypoglycemia, and titration of next insulin dose. Throughout this article, we will refer to this telemedicine intervention as '*Nurse*'.

(b) *Smartphone Application based Telecare Program (App)*: Under this program, self-titration using the smartphone app *Diabetes Pal* [11] (Fig. 1) with minimal telephone-based support from care managers for insulin initiation, was proposed. Diabetes Pal is a smartphone application that allows a diabetic patient to self-titrate their insulin doses. Patient self-titration of insulin dose based on a prescribed algorithm has been shown to be safe and efficacious in improving glycemic control [13]. The app was developed by Integrated Health Information Systems Ltd (IHiS) and has been tested for its feasibility to deliver the insulin titration algorithm in insulin-naïve patients in a pilot study recently conducted at the Singapore General Hospital [11]. Throughout this article, we will refer to this telemedicine intervention as '*App*'.

Operationalization of Competing Trial Designs
To compare a traditional RCT and SMART for evaluation of effectiveness and cost in the trial context, we implemented a microsimulation of these two trial designs run over a 12-week "study" period.
(a) *SMART Design*

The SMART design operates in two stages. At stage one, all patients were randomized with a 1:1 ratio between Nurse and App. However, at the end of stage 1 (6 weeks from the initial randomization), patients were categorized as either a responder (R = 1) or a non-responder (R = 0) based on their reduction in HbA1c value in the 6-week period. In the base-case scenario, the threshold for declaring a response was assumed to be 0.5% (i.e., the patient was considered a responder if the reduction in HbA1c from baseline is ≥ 0.5%, and a non-responder otherwise.). This threshold value was varied in sensitivity analysis. The responders to the first-stage intervention continued with the same intervention in stage 2 (weeks 6–12 of the study). However, the non-responders were re-randomized to either a switch to the intervention not tried before for the same patient or a combined intervention (App + Nurse). The non-responders were re-randomized in a 1:1 ratio between the switch option and the combination option. A schematic of the SMART design is presented in Fig. 2 (a). Because of the re-randomization at stage 2, the current SMART design offers a comparison between four embedded adaptive interventions (AIs) (see [14]), described in Table 1.

The primary outcome (Y) was the HbA1c measurement at the end of the trial, and was recorded for all patients. The expected outcomes corresponding to the four embedded adaptive interventions are denoted $E(Y)_{AI1}$, $E(Y)_{AI2}$, $E(Y)_{AI3}$ and $E(Y)_{AI4}$. Also, the average outcome corresponding to the best AI from the SMART is denoted by $\min[E(Y)_{AI1}, E(Y)_{AI2}, E(Y)_{AI3}, E(Y)_{AI4}]$ and the best AI being $\arg\min_{j \in \{1,...,4\}} E(Y)_{AIj}$; this is a key performance metric of the SMART design that we compare with the best outcome from the RCT design, in our simulation study. Note that for SMART, $E(Y)_{AI1}$, $E(Y)_{AI2}$, $E(Y)_{AI3}$ and $E(Y)_{AI4}$ need to be estimated using the inverse probability weighting method [15].

Table 1
Description of the four adaptive interventions (AI). N = Nurse, A = App and A + N = App + Nurse.

AI#	Notation	Description
AI1	$(N, N^R A^{1-R})$	Nurse at stage 1, continue with Nurse at stage 2 if responder, or switch to App at stage 2 if non-responder.
AI2	$(N, N^R (A + N)^{1-R})$	Nurse at stage 1, continue with Nurse at stage 2 if responder, or provide combination at stage 2 if non-responder.
AI3	$(A, A^R N^{1-R})$	App at stage 1, continue with App at stage 2 if responder, or switch to Nurse at stage 2 if non-responder.
AI4	$(A, A^R (A + N)^{1-R})$	App at stage 1, continue with App at stage 2 if responder, or provide combination at stage 2 if non-responder.

(b) RCT Design

The RCT design (Fig. 2 (b)) is a conventional randomization designed to test the same sequences of treatments as the SMART design with the same intermediate evaluation at 6-weeks for responsiveness to initial treatment. However, for individuals deemed not to respond to the initial treatment, the new treatment was established at the time of initial randomization rather than at 6-weeks. In other words, the four AIs are separate arms in the design, where individuals are assigned to one of the AI arms at the start of the trial. This is different from the SMART design, where the AIs are embedded. As with SMART, effectiveness and cost are assessed based on final HbA1c and cost at the trial end at 12 weeks.

Data Generation Model

Baseline HbA1c

The pilot study [11] of the Diabetes Pal smartphone app enrolled 66 insulin naïve patients with suboptimal glycemic control of HbA1c $\geq$ 7.5% despite use of 2 or more oral glucose lowering drugs. These patients were between 30 and 70 years of age. Of these 66 recruited subjects, only 63 patients had complete follow-up data. Based on the baseline HbA1c measurements of these 63 patients, a normal distribution with mean = 9.73 and SD = 1.37 was used to generate the baseline HbA1c values (Y_0) of the hypothetical patients in the simulation model. Detailed description of the model generation and algorithm may be found in the Additional file 1.

Receptiveness

In addition to the starting level of glycemic control measured by HbA1c, two additional subject characteristics were assigned at baseline: (1) receptiveness to the App (Rc_A), and (2) receptiveness to the Nurse (Rc_N). A patient is deemed "receptive" if their engagement with the specific intervention had more than a nominal impact on their tendency to improve glycemic control. This was operationalized as a binary indicator variable (1 = receptive, 0 = non-receptive). Since information on receptiveness was not collected during the pilot study, estimates from the literature were utilized. Specifically, according to Deloitte's Global Mobile Consumer Survey 2016 for UK [16], 69% of the smartphone users made standard voice calls weekly. It was assumed that the smartphone users who made standard voice calls were comfortable communicating over telephone and therefore would be receptive to receiving insulin titration information over telephone via the Nurse. The survey also reported that around 51% of the users downloaded more than five apps on their smartphones. Supported by the data in the survey, it was assumed that the users who downloaded more than five apps would be using the apps to carry out activities (other than to communicate) that involved inputting and outputting of information; it was further assumed that these people would also be receptive to using App to carry out insulin titration. We further assumed that the receptiveness to one intervention was independent of the receptiveness to the other intervention. Thus, in our microsimulation study the tendency for Rc_A and Rc_N were based on the probability of success in Bernoulli trials (receptiveness rates) 0.51 and 0.69, respectively. Furthermore, we assumed that if a subject was receptive to at least one of the two interventions, they were likely to be receptive to the combined intervention (App + Nurse). The probability of receptiveness to the combination was calculated to be 0.75.

Change in HbA1c conditional on being receptive to a received intervention

In the microsimulation, the change in HbA1c for each individual was drawn from a normal distribution corresponding to whether the subject was actually receptive or not (as assigned at baseline). As noted, receptiveness to an intervention, leading to appropriate changes in insulin dose, was assumed to be reflected in improvement in HbA1c over a 6-week period beyond random change. To generate plausible HbA1c change distributions, we stratified data from the actual trial subjects by 6-week change in HbA1c and calculated the means and standard deviations. To separate the mean reduction for receptive and non-receptive participants, we assumed an average receptiveness rate of 60% in the trial subjects and fixed the mean for non-receptive participants to be zero. Thus, the mean HbA1c reductions at week 6 and 12 for receptive participants are assumed to be 1.53 (SD = 0.71) and 0.94 (SD = 0.77), respectively.

Calculation of Trial Costs

In the microsimulation, costs were accumulated for each synthetic subject based on their interventions experienced, including the time costs of providing the interventions, monitoring, and re-randomization when needed. Because costs for the trial are almost entirely personnel time, we did not include cost of the app itself. Cost was calculated in US dollars (USD) by multiplying personnel time by the exchange rate adjusted modal wage rate for Singapore, a country with a gross domestic product per capita comparable to the US, approximately USD 57,000.

Analysis

In our simulation study, we performed a base case analysis in which the two designs were compared with the key input parameters fixed as follows:

1. receptiveness rate to the Nurse ($P(Rc_N = 1)$) = 69%,
2. receptiveness rate to the App ($P(Rc_A = 1)$) = 51%,
3. receptiveness rate to the App + Nurse ($P(Rc_{A+N} = 1)$) = 75%,
4. reduction threshold for response (δ) = 0.5%, and
5. trial size of patient cohort (n) = 100.

The Monte Carlo assessments were based on simulation size (B) = 10000. For each design, the Monte Carlo mean HbA1c estimate $\left(\tilde{y} = \frac{\sum_{b=1}^{B} \bar{y}_b}{B} \right)$ and the standard deviation $\left(s_{\tilde{y}} = \sqrt{\frac{\sum_{b=1}^{B} (\bar{y}_b - \tilde{y})^2}{(B-1)}} \right)$ were calculated where $\bar{y}_b$ is the overall and per AI mean HbA1c for the bth simulated trial. The probability of selecting the best AI as defined in *SMART design* over B simulations was calculated. In sensitivity analysis, we varied each parameter over a broad range to determine if there was any change in the sign for the difference in effectiveness or cost, large absolute changes on outcomes for both designs, or non-linear relationships between the parameters and outcomes.

Results

Base case

In the base case scenario (Table 2), the difference of the overall mean expected HbA1c and per subject cost between the SMART and the RCT design is almost negligible (both HbA1c are 8.28%; per subject cost is USD 343.32 vs USD 343.22). As the sample size increases, the overall effectiveness and cost between SMART and RCT designs are essentially equivalent (Fig. 3). This is expected since both designs are unbiased for estimating the above metrics. Although the estimated mean HbA1c by AIs are similar in both designs, the standard deviations are noticeably lower in case of the SMART design. This is true across a wide range of sample sizes, approaching zero as trial size became very large (Fig. 4). Given the underlying data generation (see Additional file 1), the true optimal AI is AI2. The SMART design is able to outperform the RCT design by having higher probability of choosing AI2 given same sample size n = 100 (48.73% vs 43.38%). The efficiency of the SMART design becomes apparent as we aim to maximize the probability of choosing AI2 by varying the sample sizes (Fig. 5). In order to have approximately 70% probability of correctly choosing AI2, SMART requires n = 500 (70.66%), whereas RCT requires n = 1700 (70.72%).

Table 2 The Monte Carlo mean (standard deviation) of the final HbA1c outcomes (Y) and cost per subject and the probability of selecting each AI as optimal (p_{optAI}) for SMART and RCT of sample size n = 100. The simulation size was 10000. According to the data generation model of the simulation study, AI2 is the truly optimal AI.

	SMART	RCT
Overall		
Y	8.23 (0.17)	8.23 (0.17)
Cost (USD)	343.32 (2.14)	343.22 (2.16)
By (embedded) AI		
Y_{AI1}	8.12 (0.28)	8.12 (0.34)
Y_{AI2}	8.04 (0.26)	8.05 (0.33)
Y_{AI3}	8.39 (0.29)	8.39 (0.34)
Y_{AI4}	8.36 (0.28)	8.36 (0.34)
p_{optAI1}	29.41%	33.77%
p_{optAI2}	48.73%	43.38%
p_{optAI3}	10.06%	10.43%
p_{optAI4}	11.80%	12.42%

Sensitivity analyses

For both trial designs the threshold value at which a subject was deemed to have been responsive at an intermediate point in the trial had an optimal value (i.e., the sensitivity curve had a U-shape) (Fig. 6). As the threshold moves away in either direction from the optimal value, the mean HbA1c for the trial subjects worsens. Under the optimal threshold, the SMART design becomes more efficient than the RCT design when the sample size is small, because the final overall mean HbA1c performs better (that is, it has a lower value). This change in threshold corresponds to a change in the relationship between the sensitivity and specificity of the interim evaluation of responsiveness. A more negative threshold results in lower sensitivity but greater specificity for responders while a higher threshold results in higher sensitivity but lower specificity.

Discussion

In this study, we examined the value of the SMART design relative to a comparable RCT design of two telemedicine interventions for insulin initiation: a largely self-contained smartphone app [11] and a nurse-based telephone consultation service. The designs were comparable in that both had the aim to evaluate the optimal sequencing of these two interventions, including the potential for combining interventions. We did this evaluation using microsimulation drawing on empirical data from a prior conventional trial. Simulation allowed us to perform sensitivity analysis of how diabetes control (as assessed by HbA1c) and trial costs were impacted by various aspects of trial design, including the operating characteristics of

the intermediate measure used in the SMART and RCT designs to continue or switch treatment. It should be noted that the RCT design used as the comparator was unconventional, involving both multiple arms and treatment switching based on interim assessment of responsiveness to the initial treatment.

While both designs provide information on the optimal sequencing of therapies, we demonstrated some notable benefits of SMART compared to RCT. First, the SMART design from the perspective of trial population, had consistently smaller variance in the mean HbA1c per AI, which was especially evident at smaller sample sizes, at approximately equivalent cost. For the same sample size, the SMART design has higher probability of identifying the best AI.

Another advantage of SMART is that the design offers the potential to personalize treatment sequences by evaluating features predictive of responsiveness by treatment order. In our present simulation study, this feature of SMART was not examined as subjects were simulated as identical with regard to all features except for responsiveness to one intervention or the other. However, there is a sizable statistical literature that offers methodologies (e.g., Q-learning) for doing such personalization as secondary analysis of SMART data [5, 17]. This aspect can be pursued in simulations as an important future work.

In sensitivity analysis, the observed benefits were robust. However, we did note that the value of both designs depended on the threshold value for defining response to treatment at the end of first stage. Average HbA1c control for trial subjects was optimal at an intermediate threshold value: too low and subjects who were unresponsive to their initial treatment were incorrectly maintained on an ineffective therapy; too high and subjects who were responsive to initial therapy would be incorrectly switched from an effective therapy. This suggests that the sensitivity and specificity of the threshold value can be important parameters to consider in SMART design and that the value of the design can be much diminished if the first stage evaluation does not have good operating characteristics.

Most clinical trials aim to conduct formal hypothesis tests in order to determine the superior interventions. However, in case of telemedicine, it may often be of more interest to find out if a cheaper or less burdensome intervention (e.g. App) is non-inferior to an established but more expensive intervention (e.g. Nurse). Such non-inferiority testing methodologies have been applied to conventional RCTs for many years [18]. Very recently, such non-inferiority testing methods [19] along with free web-based software [20], have also been developed in the SMART design context. Availability of such methodology and software tools brings SMARTs to an even playing field as RCTs, in terms of flexibility of hypothesis testing and data analysis. We have not considered non-inferiority testing in the current manuscript.

The primary goal of SMART is to learn – through *within-patient adaptation* of interventions over stages – an optimal strategy that can benefit future patients beyond the trial, not the trial participants *per se*. As such, it does not allow *between-patient adaptation* of interventions within the trial, because the randomization probabilities in a SMART are pre-specified. This fixed allocation scheme in a SMART design (as in conventional RCT) is motivated by the aim to maximize statistical power in order to maximize the scientific information gained from the trial. However, there are settings (e.g., implementation studies) where there is urgent need to translate emerging evidence from ongoing trials

into practice, including the remainder of the trial participants, in order to maximize the benefit to the overall population of interest [20]. This need can be accommodated in both a SMART and an RCT through the machinery of response-adaptive allocation. Such an adaptive SMART or adaptive RCT design would allow modification of the randomization probabilities based on observed outcome data, favoring the treatment sequences that empirically look better (even though not statistically significant), at pre-set interim times during the trial [6, 21, 22]. For simplicity, we chose not to consider such a response-adaptive SMART or RCT in our current simulation study. However, we feel that such designs can potentially be even more attractive in the telemedicine context, optimizing welfare of trial participants while also finding optimal care strategies for future patients. We view more in-depth study of such designs in the telemedicine arena as an important future work.

Conclusion

In light of increasingly complex care management questions, new trial designs have been offered to improve the range of useful inferences that can be derived from clinical trials. SMART is one example that is particularly suited to evaluating the efficacy of different sequences of treatment options. To make better use of SMART, it is important to understand the advantages and disadvantages of SMART relative to a conventional design. This study illustrates the advantages of the SMART design over a comparable RCT for evaluating sequences of therapies. We note that the value of a SMART depends on the accuracy of the intermediate measure of responsiveness as well as the burden and cost of re-randomization.

Abbreviations

SMART: Sequential Multiple Assignment Randomized Trial

RCT: Randomized Controlled Trial

AI: Adaptive Interventions

A: App, a smartphone application, *Diabetes Pal*

N: Nurse, a nurse-based telemedicine strategy

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

References

1. Lei, H., et al., *A "SMART" design for building individualized treatment sequences*. Annu Rev Clin Psychol, 2012. **8**: p. 21–48.
2. Murphy, S.A., *An experimental design for the development of adaptive treatment strategies*. Statistics in Medicine, 2005. **24**(10): p. 1455–1481.
3. 2017 *ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA Guideline for the Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines*. Journal of the American College of Cardiology, 2018. **71**(19): p. e127-e248.
4. Lavori, P.W. and R. Dawson, *Adaptive Treatment Strategies in Chronic Disease*. Annual Review of Medicine, 2008. **59**(1): p. 443–453.
5. Chakraborty, B. and E.E.M. Moodie, *Statistical methods for dynamic treatment regimes: reinforcement learning, causal inference, and personalized medicine*. 2013, New York, NY: Springer.
6. Thall, P.F., H.-G. Sung, and E.H. Estey, *Selecting Therapeutic Strategies Based on Efficacy and Death in Multicourse Clinical Trials*. Journal of the American Statistical Association, 2002. **97**(457): p. 29–39.

7. Kumar, S., et al., *Mobile health technology evaluation: the mHealth evidence workshop.* American journal of preventive medicine U6 - ctx_ver = Z39.88-2004&ctx_enc = info%3Aofi%2Fenc%3AUTF-8&rfr_id = info%3Asid%2Fsummon.serialssolutions.com&rft_val_fmt = info%3Aofi%2Ffmt%3Akev%3Amtx%3Ajournal&rft.genre = article&rft.atitle = Mobile + health + technology + evaluation%3A + the + mHealth + evidence + workshop&rft.jtitle = American + journal + of + preventive + medicine&rft.date = 2013-08-01&rft.eissn = 1873–2607&rft.volume = 45&rft.issue = 2&rft.spage = 228&rft.epage = 236&rft_id = info:doi/10.1016%2Fj.amepre.2013.03.017&paramdict = en-US U7 - Journal Article, 2013. **45**(2): p. 228–236.

8. Craig, J. and V. Patterson, *Introduction to the practice of telemedicine.* J Telemed Telecare, 2005. **11**(1): p. 3–9.

9. Sood, S., et al., *What is telemedicine? A collection of 104 peer-reviewed perspectives and theoretical underpinnings.* Telemed J E Health, 2007. **13**(5): p. 573–90.

10. Suter, P., W.N. Suter, and D. Johnston, *Theory-based telehealth and patient empowerment.* Popul Health Manag, 2011. **14**(2): p. 87–92.

11. Bee, Y.M., et al., *A Smartphone Application to Deliver a Treat-to-Target Insulin Titration Algorithm in Insulin-Naive Patients With Type 2 Diabetes: A Pilot Randomized Controlled Trial.* Diabetes care, 2016. **39**(10): p. e174-e176.

12. R Core Team, *R: A language and environment for statistical computing. R Foundation for Statistical Computing.* 2020: Vienna, Austria.

13. Khunti, K., M.J. Davies, and S. Kalra, *Self-titration of insulin in the management of people with type 2 diabetes: a practical solution to improve management in primary care.* Diabetes Obes Metab, 2013. **15**(8): p. 690–700.

14. Nahum-Shani, I., et al., *Experimental design and primary data analysis methods for comparing adaptive interventions.* Psychol Methods, 2012. **17**(4): p. 457–477.

15. Robins, J.M., M.A. Hernan, and B. Brumback, *Marginal structural models and causal inference in epidemiology.* 2000, LWW.

16. Deloitte, *Global Mobile Consumer Survey 2016: The UK Cut.* 20, Deloitte: London, UK.

17. Kosorok, M. and E. Moodie, *Adaptive Treatment Strategies in Practice.* Adaptive Treatment Strategies in Practice.

18. D'Agostino, R.B., Sr., J.M. Massaro, and L.M. Sullivan, *Non-inferiority trials: design concepts and issues - the encounters of academic consultants in statistics.* Stat Med, 2003. **22**(2): p. 169–86.

19. Ghosh P, et al., *Non-inferiority and equivalence tests in sequential, multiple assignment, randomized trials (SMARTs).* Psychological Methods, 2019.

20. P, G., *Shiny App.* 2019.

21. Thall, P.F. and J.K. Wathen, *Covariate-adjusted adaptive randomization in a sarcoma trial with multi-stage treatments.* Stat Med, 2005. **24**(13): p. 1947–64.

22. Cheung, Y.K., B. Chakraborty, and K.W. Davidson, *Sequential multiple assignment randomized trial (SMART) with adaptive randomization for quality improvement in depression treatment program.* Biometrics, 2015. **71**(2): p. 450–9.
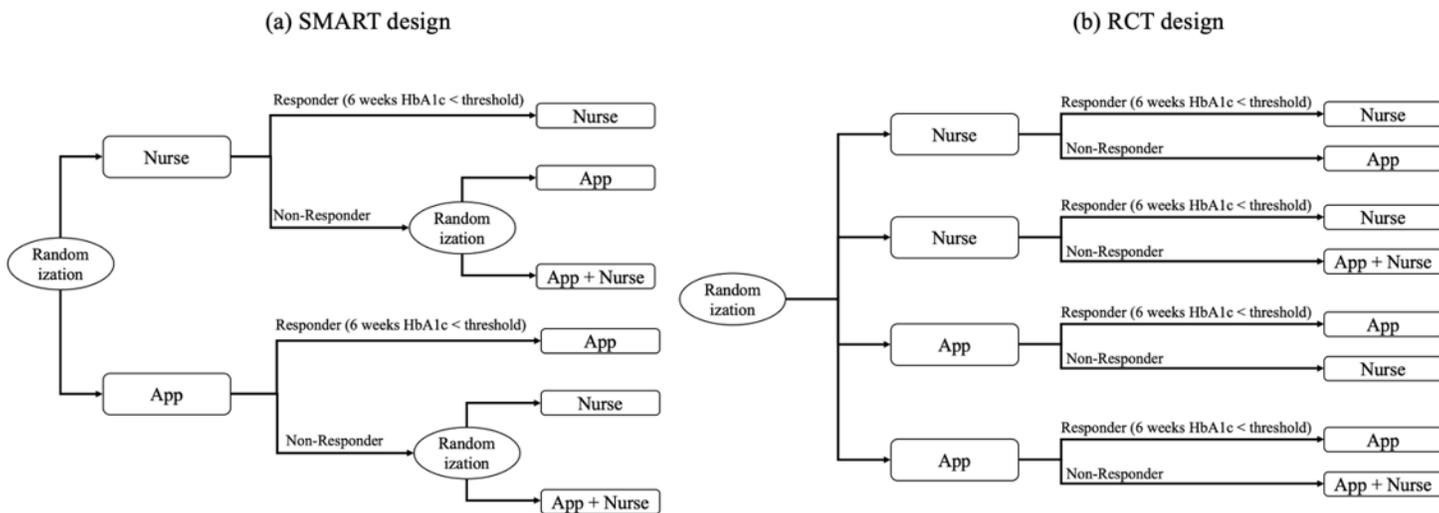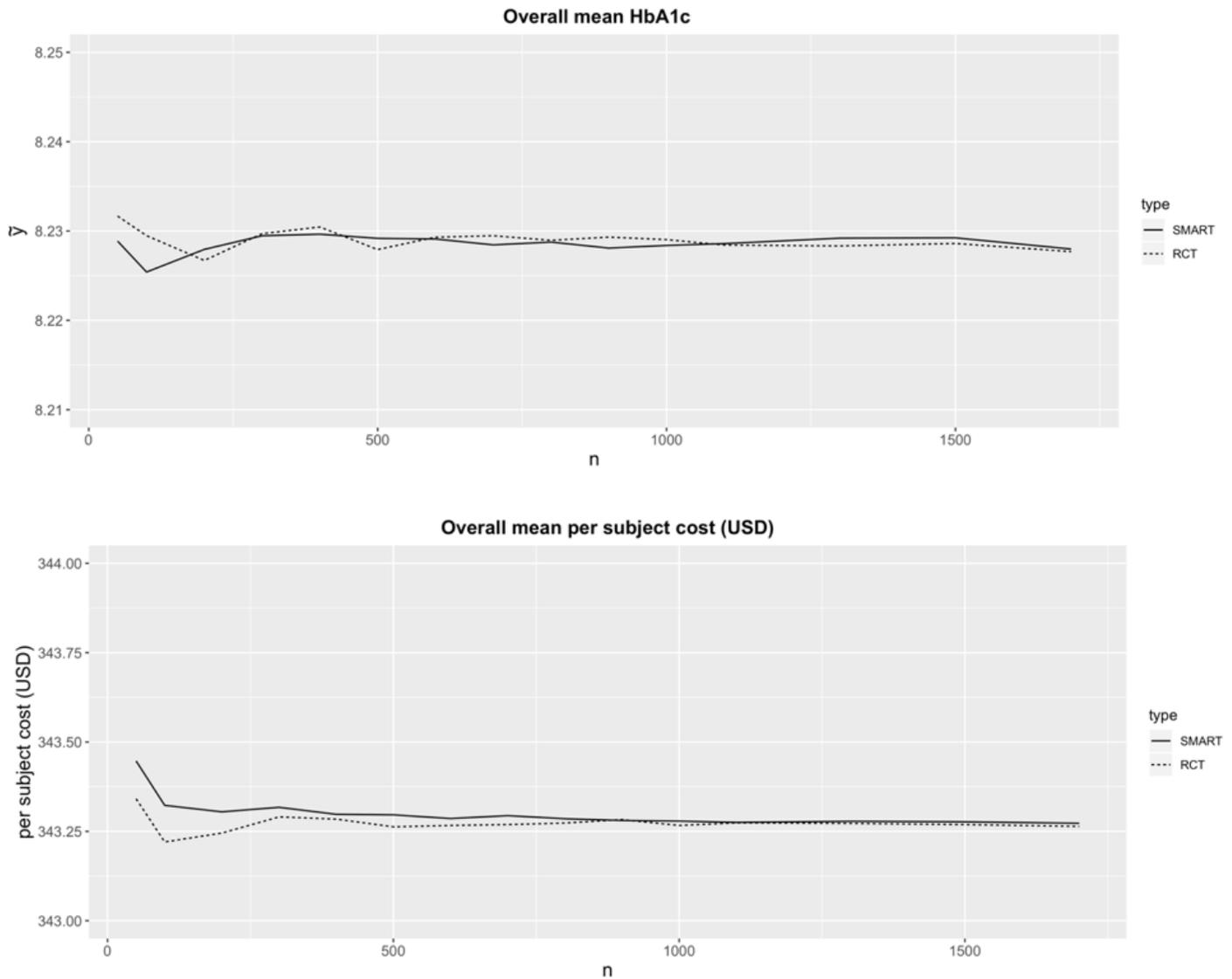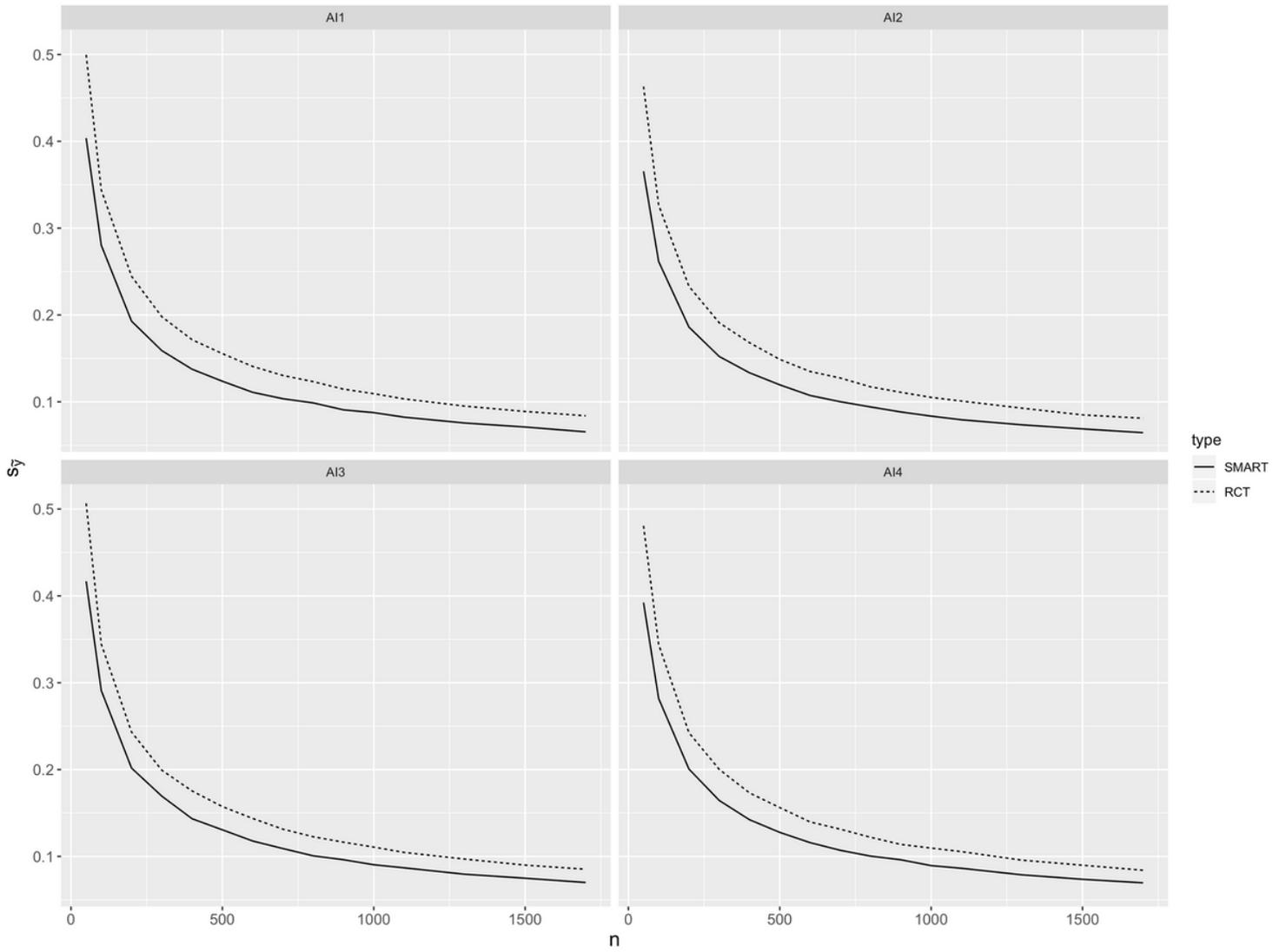
Figures

Figure 1

Diabetes pal

Figure 2

The (a) SMART and (b) RCT design in comparison. Each intervention component (Nurse, App, App + Nurse) is for 6 weeks long. (a) Randomisation happens at baseline for stage 1 intervention and at week 6 for stage 2 intervention, given stage 1 treatment and response. The four AIs are embedded in the design as a result. (b) Randomisation happens at baseline only, so the four arms correspond to the four AIs.

Figure 3

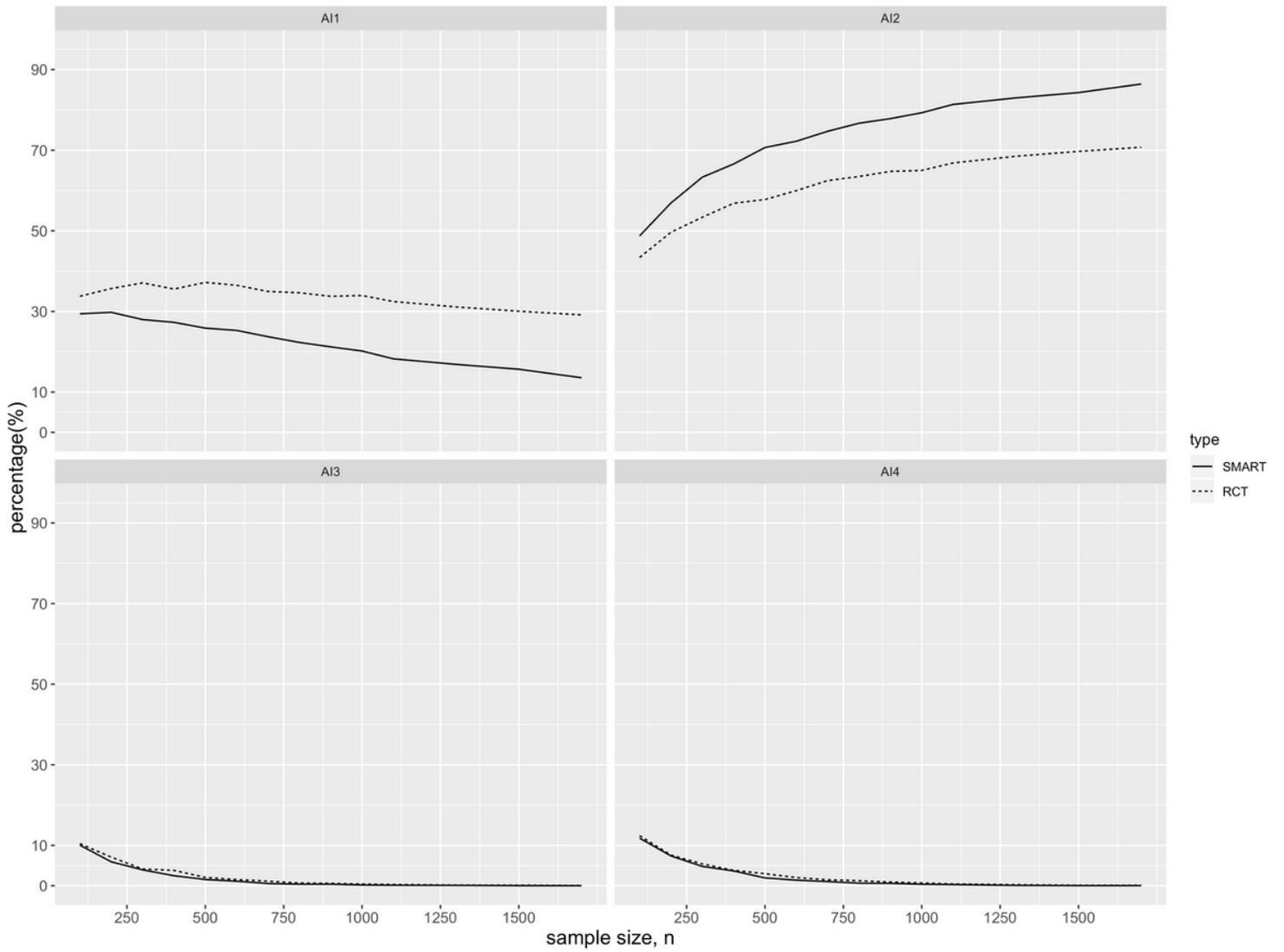The Monte Carlo means of the final outcome HbA1c and per subject cost (USD) for different sample sizes.
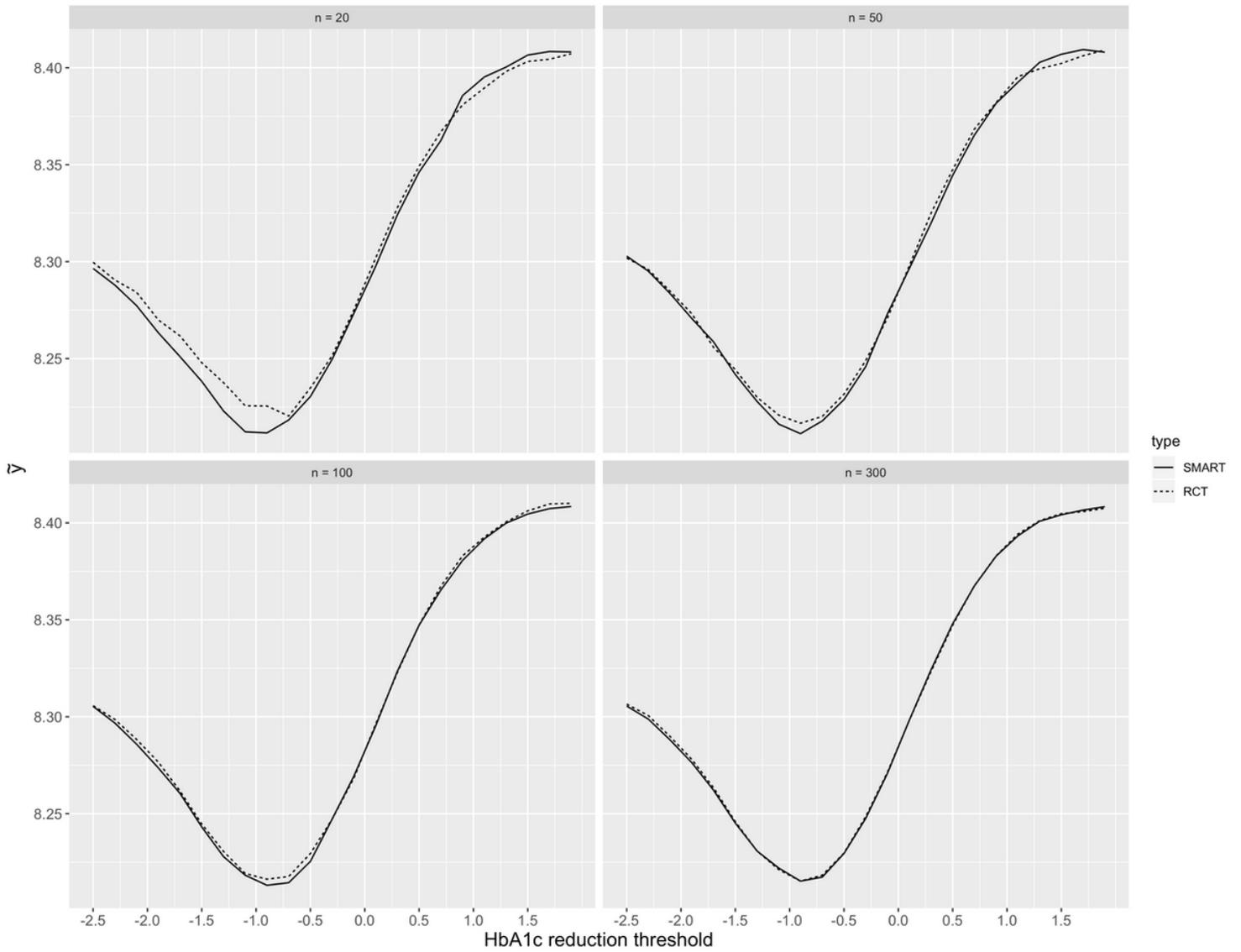
Figure 4

The standard deviations of the Monte Carlo means of the final outcome HbA1c by AIs for different sample sizes.

Percentage of times AI_j identified as best

Figure 5

The number of times out of B = 10000 simulations, the SMART and RCT trial identify the j-th AI (j=1, …, 4) as the best AI across a wide range of sample sizes. According to the data generation model of the simulation study, AI2 is the truly optimal AI.

Figure 6

The overall mean expected HbA1c from 10000 simulated SMART and RCT trials for sample size n = 20, 50, 100, 300 varied across a wide range of reduction threshold values.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile1.docx