British Journal of Cancer www.nature.com/bjc

REVIEW ARTICLE

Clinical Studies

Use of Sequential Multiple Assignment Randomized Trials (SMARTs) in oncology: systematic review of published studies

Giulia Lorenzoni[1,3], Elisabetta Petracci[2,3], Emanuela Scarpi[2], Ileana Baldi[1], Dario Gregori[1]✉ and Oriana Nanni[2]

Sequential multiple assignments randomized trials (SMARTs) are a type of experimental design where patients may be randomised multiple times according to pre-specified decision rules. The present work investigates the state-of-the-art of SMART designs in oncology, focusing on the discrepancy between the available methodological approaches in the statistical literature and the procedures applied within cancer clinical trials. A systematic review was conducted, searching PubMed, Embase and CENTRAL for protocols or reports of results of SMART designs and registrations of SMART designs in clinical trial registries applied to solid tumour research. After title/abstract and full-text screening, 33 records were included. Fifteen were reports of trials' results, four were trials' protocols and fourteen were trials' registrations. The study design was defined as SMART by only one out of fifteen trial reports. Conversely, 13 of 18 study protocols and trial registrations defined the study design SMART. Furthermore, most of the records considered each stage separately in the analysis, without considering treatment regimens embedded in the trial. SMART designs in oncology are still limited. Study powering and analysis is mainly based on statistical approaches traditionally used in single-stage parallel trial designs. Formal reporting guidelines for SMART designs are needed.

British Journal of Cancer; https://doi.org/10.1038/s41416-022-02110-z

INTRODUCTION

Dynamic Treatment Regimens (DTRs), also known as adaptive treatment strategies or adaptive interventions, are a set of sequential decision rules, each one corresponding to a key decision point in the patient's history [1]. Each rule establishes the treatment for the patient among the available treatment options according to the information collected until then.

The DTR represents a formalisation of the multi-stage and dynamic decision process followed by clinicians in their everyday clinical practice. The final aim of the decision process is to tailor the treatment to the patients' characteristics and clinical history, which is the key concept of precision medicine. In this sense, identifying the optimal DTR would be a way to put evidence-based precision medicine into practice, especially in chronic disease management [2], which is one of the most suitable clinical settings for DTRs. Particularly, cancer research is a promising field of application of SMART designs. Cancer is a chronic disease that requires treatment at multiple stages, according to each patient's characteristics and clinical status [3].

However, providing evidence-based DTRs poses relevant methodological challenges to study design and DTRs' effect estimation. The types of study commonly used for testing and comparing DTRs include observational studies, one-time randomised trials that randomise patients only once to the whole DTR, and sequential multiple assignment randomized trials (SMARTs) [2]. SMART designs randomise patients at each decision point considering information collected on the patient so far. They are of growing interest in the scientific community, but their use is not well-established yet.

The main difficulty in implementing trials to study DTRs is that there are still several open questions about sample size calculation and identification of the most appropriate method for data analysis [4]. In oncology, as in many other chronic conditions, it is common that the patient receives a frontline treatment followed by subsequent treatments adaptively chosen by the clinician. Consequently, the patient's survival depends not only on the frontline treatment but also on the entire treatment strategy. However, the literature seems to be still dominated by trials investigating a single line or stage of the patient clinical history, ignoring previous or subsequent therapies, potentially leading to misleading results [5].

The present work aims to investigate the state-of-the-art of SMART designs in oncology, focusing on the statistical methods used for the sample size computation and data analysis within cancer clinical trials on solid tumours.

METHODS

A systematic review was done. The review was conducted according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [6].

[1]Unit of Biostatistics, Epidemiology and Public Health, Department of Cardiac Thoracic Vascular Sciences and Public Health, University of Padova, Padova, Italy. [2]Unit of Biostatistics and Clinical Trials, IRCCS Istituto Romagnolo per lo Studio dei Tumori (IRST) "Dino Amadori", Meldola, Italy. [3]These authors jointly supervised this work: Giulia Lorenzoni, Elisabetta Petracci. ✉email: dario.gregori@unipd.it

Received: 20 July 2022 Revised: 5 December 2022 Accepted: 7 December 2022
Published online: 26 December 2022

Information sources and search strategy

The bibliographic search was performed on PubMed, Embase and CENTRAL (Cochrane Trial Registry), without date of publication restrictions. The search string is reported in Table S1 (Supplementary Material).

Eligibility criteria and selection process

Published protocols or results of SMART designs and registrations of SMART designs in clinical trial registries were considered eligible. To be included, the SMART design should be applied to solid tumour research, without restrictions on the intervention type.

The criterion to identify SMART designs was the presence of ≥2 stages in which patients were re-randomised to subsequent treatments according to a set of pre-specified decision rules based on patients' characteristics and treatment history [7].

The study selection was done using the COVIDENCE software [8]. The title/abstract and full-text screening was performed by two independent reviewers (GL and EP). A third independent reviewer (ES) was in charge of solving disagreements.

Conference proceedings, book chapters, systematic reviews and metanalysis were excluded, but they were checked for eligible papers. Papers in the English language were considered.

Data extraction

Information on three domains of interest was considered, i.e. study characteristics, study design and study analysis. Study characteristics included publication year, setting, funding, trial registration (if any), the definition of the study design as SMART, and if the study presented a reanalysis of the original study data. The study design information included the number of stages, the type of intervention administered at each stage, the decision rules employed, the study objectives and endpoints and sample size reporting and calculation information. The study analysis domain included the methods used for data analysis and if specific data analysis techniques were used to account for the adaptive treatment resulting from the multiple sequential assignments. For protocols, such information was extracted from the statistical analysis plan.

A restricted subset of items was employed to extract data from trials' registrations records to allow a minimum dataset for all trials' registrations included in the review. The item selection depended on the fact that the detail of information reported slightly changed according to the trial registry type. In most cases, the statistical analysis plan was missing.

Study characteristics were reported for descriptive purposes. Study design information was chosen according to the key SMART designs' components, e.g. the number of stages, decision rules. Finally, information on study analysis techniques allowed for answering the main research question of the present work, i.e. to describe the statistical methods used within SMART designs to identify potential discrepancies between the available methodological approaches in the statistical literature and the procedures applied.

The data extraction tool was based on an Excel file.

Risk of bias assessment

The Cochrane risk-of-bias tool for randomised trials (RoB 2) [9] was used to assess the risk of bias in the included studies (articles and protocols). For studies included twice in the review [10–13], the assessment of the risk of bias was performed only once.

RESULTS

Search results

The search of the bibliographic databases resulted in the inclusion of 14,586 records (Fig. 1). The last search was performed on 9 September 2021.

After duplicate removal, title/abstract screening was performed, resulting in 823 included records which underwent full-text screening. After the full-text screening, 33 results were included in the present systematic review. Fifteen were reports of trials' results [10–24], four were trials' protocols [25–28] and fourteen were trials' registrations.

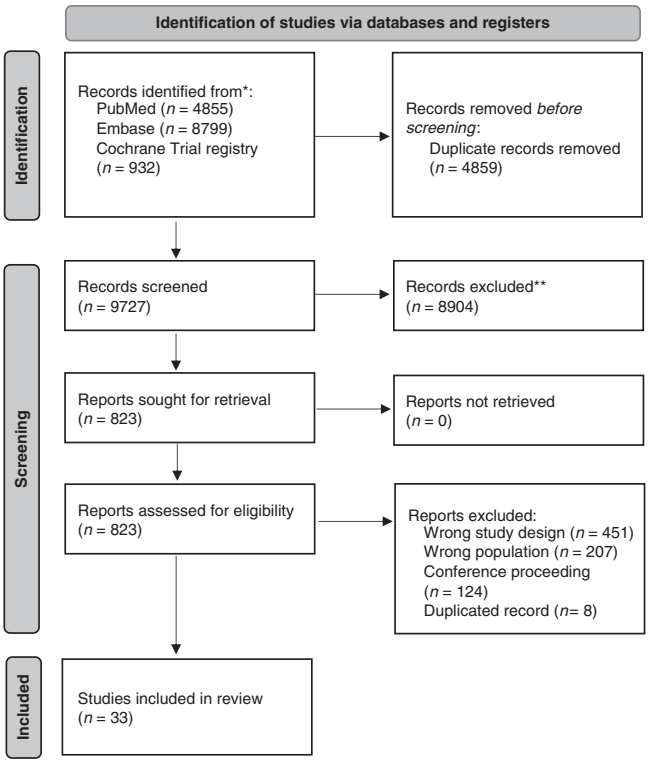

Fig. 1 Study flow-chart. PRISMA flow-chart showing the study selection process.

Among the included records, there was a match between three trials' protocols [26–28] and the corresponding trials' registrations, and between two reports of trials' results [16, 22] and the corresponding trials' registrations.

Among trials' results reports and protocols, nine were published in oncology journals [11, 12, 14, 15, 18, 19, 22–24], three in experimental and research medicine journals [25–27], two in internal and general medicine journals [10, 16]. The other five records were published in specialised journals in other areas of medicine, including clinical neurology [17], respiratory system [20] and peripheral vascular diseases [21]. One study was published in a nursing journal [28], and only one study was published in a statistics & probability journal [13].

All studies were found to present with some concerns at risk of bias assessment (Table S2, Supplementary Material), except for that of Marshall et al. [21].

Trials' results

Fifteen studies presenting trials' results were included in the present work. Table S3, Supplementary material, presents the detailed characteristics of the studies included.

Eight trials were located in the EU and 5 in North America. Thirteen out of fifteen were multicenter, and half (8 out of 15) received public or private funding. The first study was published in 1992. Six trials were published between 2010–2021, three between 2000–2009 and another six in the period 1990–1999.

Two studies [13, 22] presented a reanalysis of previously published data; for what concerns that of Wang et al. [13], the study presenting the first analysis of the data was included in this review [12], whereas for the one by Petracci et al. [22], it was included the one reporting the reanalysis.

Furthermore, two studies presented the same trial's short- and long-term results [10, 11].

All the trials tested chemo/radio/hormone therapy for cancer treatment, including lung cancer, neuroblastoma, glioblastoma, pancreatic cancer, breast cancer, prostate cancer, colorectal cancer and recurrent venous thrombosis in solid tumours.

Interestingly, only one study out of the fifteen included was reported to have a SMART design [13]. All the trials were characterised by a two-stage design (Table 1 for detailed study design). The decision rule was most frequently based on the response to first-stage treatment. The objectives reported by most of the studies identified were to compare first and second-stage treatments separately or only first or second-stage treatments, except for Petracci et al. [22], Thall et al. [12] and Wang et al. [13]. The authors of these studies explicitly declared in the manuscript that the study's objective was to identify the best treatment regimen resulting from the multiple assignments.

Eight studies did not report sample size calculation. Those reporting sample size calculations did not take into account the multiple assignments in the sample size estimation. Generally, the sample size was provided for each stage, or the powering of the study was made on one of the two stages and inflating according to the expected proportion of subjects entering the second randomisation. Of notice, Tummarello et al. [24] declared that the number of people entering the second randomisation was too small to allow groups' comparison. Marshall et al. [21] and Bianchi et al. [14] underwent premature closure because of the low recruitment rate.

Regarding the data analysis (Table 2 for study details), the approaches most frequently used were the Kaplan–Meier method and the Cox Proportional Hazard model since most trials considered a time-to-event endpoint (overall or progression-free survival). Matthay et al. [10, 11] used such analysis approaches to compare the treatment regimens resulting from the two-stage randomisation among subjects entering the second randomisation. In all other trials, separate analyses of first and second-stage treatments were carried out, except for Petracci et al. [22], Thall et al. [12] and Wang et al. [13], which were interested in identifying the best treatment regimen. These authors adopted three different strategies of analysis to estimate the treatment effect taking into account patients' baseline characteristics and outcome history throughout the trial. The analysis of Petracci et al. [22] involved the estimation of Inverse Probability of Censoring Weighting (IPCW) to account for selection bias resulting from patients' selection in the second stage. Thall et al. [12] used a conditional logistic regression approach. Wang et al. [13] proposed the estimation of Inverse Probability Treatment Weighting (IPTW) for the reanalysis of Thall et al. [12].

Trials' protocols

The review included four trial protocols [25–28]. They were all located in the USA and published after 2009. Three [26–28] out of four corresponded to trials' registrations included in the present review. All the study's protocols received funding, and two were multicenter (Table S3 for trials' protocols characteristics).

No trials were aimed at testing chemo/radiotherapy for cancer treatment. Two tested interventions to reduce cancer symptoms in patients with different types of solid tumours [28] and breast cancer [27]. One tested pharmacological treatment for depression in melanoma patients undergoing IFN-alpha therapy [25]. Finally, that of Fu et al. [26] was aimed at lung cancer prevention through a smoking cessation programme.

In all four protocols included, the design was defined to be SMART. They were all based on two stages, and the decision rule was based on the response to the first-stage treatment (Table 1). All the protocols were declared to be aimed at identifying the optimal treatment strategy. However, it is worth pointing out that only one protocol presented the identification of the optimal treatment strategy as the study's primary objective [25].

All protocols reported the sample size calculation, but the power analysis was based only on one of the two stages in three out of four records. Only Auyeung et al. [25] proposed an approach accounting for the two-stage design. Not least, Fu et al. [26], in the last trial's update published within the trial registration, declared that a sample size reassessment was done to account for the low enrolment rate.

For data analysis (Table 2), all protocols planned to use traditional statistical tests and regression-based analyses to compare first and second-stage treatment separately. Furthermore, Sikorsii et al. [28], Kelleher et al. [27] and Auyeung et al. [25] proposed three different analysis approaches to identify the optimal treatment strategy. Auyeung et al. [25] proposed using marginal mean models to estimate the mean outcome for each regimen. Sikorsii et al. [28] declared that the optimal intervention sequence will be identified through Q-learning algorithm, including two Q functions considering patients and their caregivers' baseline characteristics and history through the two stages. Also, Kelleher et al. [27] planned the use of the Q-learning algorithm and value search estimation. No technical details about models' estimations were provided.

Trials' registrations

Fourteen trials' registrations were included in the review. Five referred to already included trials' results [16, 22] and protocols [26–28]. All registrations were made after 2008, twelve were retrieved on clinicaltrials.gov, one from australianclinicaltrials.gov and one from the Clinical Trials Peruvian Registry. Half of the studies were located in North America.

Only five out of 14 trials (36%) were aimed at testing cancer chemo/radio treatments on overall survival or disease-free survival of patients with pancreatic cancer (3 registrations), colorectal cancer (1 registration) and neuroblastoma (1 registration). Six trials tested treatments for cancer and cancer treatment symptoms, such as fatigue, pain, sensory symptoms, depression, anxiety and quality of life, in patients with breast cancer or solid tumours and their caregivers. One trial's registration was aimed at improving the management of cardiovascular comorbidities in cancer patients, and another one at testing interventions for COVID-19 prevention and treatment in cancer patients. Finally, one trial was aimed at cancer prevention (lung cancer), through a programme for smoking cessation, corresponding to the registration of the trial protocol published by Fu et al. [26].

Interestingly, all but five registrations referred to the study design as a SMART one. All designs were two-stage based, except for two studies. One included three stages, but only one decision-rule-based randomisation was specified (from the second to the third stage), while in the other trial, the number of stages depended on the patients' COVID-19 status (no exposure, exposure to COVID-19, moderate or severe COVID-19 infection).

No information is reported regarding sample size calculation and data analysis because the statistical analysis plan was not available in almost all trials' registrations.

Detailed characteristics of each one of the trials' registrations included in the review are reported in Table S4, Supplementary Material.

DISCUSSION

One of the most relevant findings of the present systematic review is the low number of studies retrieved. Such a low number of records suggests that the use of SMART designs in oncology is still limited, even though the advent of SMART designs offers new opportunities to develop evidence-based personalised treatment regimens, especially in cancer research [7]. Such findings could be related to the fact that they pose relevant methodological challenges to the sample size and treatment effect estimation and that there is still limited dissemination and perhaps understanding of the methods in the SMART research area. Unsurprisingly, most of the studies employed traditional techniques for study powering and analysis, considering each stage separately instead of comparing DTRs embedded in the trial, maybe because

Table 1. Study design information on reports of trial's results and protocols.

First author	Number of stages	Interventions per stage	Decision rule	Study objectives	Study endpoints (if reported in the article, distinction is made between primary and secondary endpoint)	Sample size calculation reporting	Does the sample size calculation take into account the multiple stages of randomisation?
Auyeung et al. [25] (Protocol)	Two	I stage: escitalopram Vs. methylphenidate II stage: switch to the alternative drug Vs. initial drug + alternative drug	Score on Hamilton Depression Rating Scale (HAM-D)	(1) To identify the optimal dynamic treatment strategy (2) To compare first-stage treatments (3) To compare second-stage treatments	Primary endpoint: Adherence to 12 weeks of IFN-alpha treatment Overall drop-out rate Secondary endpoints: Number of IFN-alpha treatments tolerated at first stage Overall drop-out rate per stage Symptomps of depressed mood, anxiety, irritability Symptomps of fatigue, anorexia, altered sleep, psychomotor slowing Tertiary endpoints: Number of IFN-alpha treatments tolerated in those undergoing switch Vs. augmented therapy Overall drop-out rate in those undergoing switch Vs. augmented therapy	Yes	Yes. Simulations of the power achieved by the study were made for both (1) and (2) study's objectives. For study's objective (3), the minimal detectable standardized effect size was reported according to different remission rates.
Bianchi et al. [14] (Trials' results)	Two	I stage: maintenance Vs. suspension of androgen deprivation therapy + docetaxel II stage: continuing docetaxel till progression or 10-cycle completion Vs. docetaxel interruption unitl PSA rose by 50% or disease progression	PSA response (>50% reduction from baseline)	(1) To demonstrate non-inferiority of androgen deprivation therapy suspension (first stage) (2) To compare second-stage treatments	Primary endpoint: Overall survival Secondary endpoints: Time to biochemical progression Clinical progression Death from any cause Time to metatsatic progression PSA response Toxicity Quality of life	Yes	No (sample size based on upfront comparision). Power analysis reported for the second stage according to the expected proportion of subjects undergoing the second stage
Fisher et al. [15] (Trials' results)	Two	I stage: tamoxifen Vs. placebo II stage: additional 5 years of tamoxifen Vs. placebo	Complete initially assigned 5-year tamoxifen treatment No therapy discontinuation No tumour recurrence/ secondary primary cancer	To compare second-stage treatments	Overall survival Disease-free survival	No	

Table 1. continued

First author	Number of stages	Interventions per stage	Decision rule	Study objectives	Study endpoints (if reported in the article, distinction is made between primary and secondary endpoint)	Sample size calculation reporting	Does the sample size calculation take into account the multiple stages of randomisation?
Fu et al. [26] (Protocol)	Two	I stage: 4 Vs. 8 weeks tobacco longitudinal care II stage: continuing tobacco longitudinal care Vs. tobacco longitudinal care + medication therapy management (incomplete responders) continuing tobacco longitudinal care Vs. less frequet tobacco longitudinal care (complete responders)	Response to treatment (complete/ incomplete)	(1) To compare second-stage treatments among incomplete responders (2) To examine the role of timing of assessment in treatment response (3) To compare second-stage treatments in complete responders (4) To examine the role of smoking in the 7 days prior to respond assessment in incomplete responders (5) Role of lung cancer screening results in smoking abstinence moderation	Primary endpoint: Prolonged smoking abstinence Secondary endpoints: Other smoking outcomes Treatment utilisation and satisfaction Lung cancer screening results Perception of cancer risk	Yes	The study is powered for the second stage. The final sample size is based on the expected incomplete responders in the first stage
Hammel et al. [16] (Trials' results)	Two	I stage: induction with gemcitabine Vs. gemcitabine + erlotinib II stage: chemotherapy Vs. chemoradiotherapy	Controlled tumour and WHO performance status ≤ 2	(1) To compare second-stage treatments (from the date of first randomisation) (2) To compare first-stage treatments	Primary endpoint: Overall survival Secondary endpoints: Progression-free survival	Yes	No. The sample size calculation was made on the effect size for chemoradiotherapy (second-stage treatment) considering the overall survival from the date of first randomisation. The final sample size is based on the expected proportion of patients ineligible to the second randomisation
Hovey et al. [17] (Trials' results)	Two	I stage: bevacizumab Vs. bevacizumab+carboplatin II stage: continue bevacizumab Vs. cease bevacizumab	Progression	To compare second-stage treatments	Primary endpoint: Progression-free survival Secondary endpoints: Response rate Overall survival Health-related quality of life Cognitive function Corticosteroid dose Toxicity	Yes	No
Joss et al. [18] (Trials results)	Two	I stage: randomisation to one of three induction regimens II stage: maintenance chemotherapy + radiotherapy Vs. maintenance chemotherapy alone	Complete or partial remission	(1) To compare first-stage treatments (2) To compare second-stage treatments	Survival	No	

Table 1. continued

First author	Number of stages	Interventions per stage	Decision rule	Study objectives	Study endpoints (if reported in the article, distinction is made between primary and secondary endpoint)	Sample size calculation reporting	Does the sample size calculation take into account the multiple stages of randomisation?
Kelleher et al. [27] (Protocol)	Two	I stage: brief Vs. full behavioural pain intervention II stage: maintenance dose Vs. no intervention (if pain reduction ≥ 30%) increased dose Vs. maintenance dose (if pain reduction > 30%)	Pain rating	(1) To compare first-stage treatments (2) Comparison between intervention dose sequences (3) Identification of patients' charactersitics that moderate responses at each stage (4) Cost-effectiveness analysis	Primary ednpoint: Percent reduction in pain	Yes	No (sample size based on upfront comparision)
Kubota et al. [19] (Trials' results)	Two	I stage: randomisation to 3 chemotherapy regimens II stage: chest irradiation Vs. no chest irradiation	Stage III disease	(1) To compare first-stage treatments (2) To evaluate the role of chest irradiation combined with chemotherapy	Survival Response rate Response duration Progression	Yes	No (sample size based on upfront comparision)
Lebeau et al. [20] (Trials' results)	Two	I stage: aspirin Vs. no aspirin (all patients received six courses of combined chemotherapy) II stage: six more courses of chemotherapy Vs. no chemotherapy until relapse	Complete response	To compare second-stage treatments	Survival Tumour response	No	
Marshall et al. [21] (Trials' results)	Two	I stage: dalteparin Vs. rivaroxaban II stage: rivaroxaban Vs. placebo	Residual deep vein thrombosis at compression ultrasound or had suffered of pulmonary embolism	To compare second-stage treatments	Primary endpoint: VTE recurrence at 12 months after first randomisation Secondary endpoints: Major bleeding Clinically relevant non-major bleeding Overall survival VTE recurrence at 12 months for the subgroup of patients with no residual deep vein thrombosis	Yes	No (sample size based on upfront comparision). It was reported that the sample size was estimated taking into account that it was large enough to allow sufficient numbers of patients to be enrolled in the second stage
Matthay et al. (a) [10] (Trials' results)	Two	I stage: transplantation Vs. continuation chemotherapy II stage: 13-cis-retinoic acid Vs. no further therapy	No disease progression	(1) To compare first-stage treatments (2) To compare second-stage treatments	Event-free survival	No	
Matthay et al. (b) [11] (Trials' results)	Two	I stage: transplantation Vs. continuation chemotherapy II stage: 13-cis-retinoic acid Vs. no further therapy	No disease progression	(1) To compare first-stage treatments (2) To compare second-stage treatments (3) Identification of prognostic factors	Event-free survival	No	

Table 1. continued

First author	Number of stages	Interventions per stage	Decision rule	Study objectives	Study endpoints (if reported in the article, distinction is made between primary and secondary endpoint)	Sample size calculation reporting	Does the sample size calculation take into account the multiple stages of randomisation?
Petracci et al. [22] (Trials' results)	Two	I stage: Chemotherapy and bevacizumab followed by chemotherapy alone (Arm A) Vs. chemotherapy alone followed by chemotherapy and bevacizumab (Arm B) II stage: Chemotherapy Vs. Chemotherapy and cetuximab for Arm A; Chemotherapy + Bevacizumab Vs. Chemotherapy + Bevacizumab + Cetuximab for Arm B	Disease progression in wild-type KRAS patients	To assess the overall effect of bevacizumab	Progression-free survival	No	
Sculier et al. [23] (Trials' results)	Two	I stage: doxorubicin Vs. epirubicin in induction chemotherapy (first series of patients); epirubicin 60 Vs. epirubicin 90 in induction chemotherapy (second series of patients) II stage: no chemotherapy Vs. maintenance chemotherapy	Complete or partial response	(1) To compare second-stage treatments (2) To compare first-stage treatments	Survival/Disease progression	Yes	No (sample size based on the second stage)
Sikorskii et al. [28] (Protocol)	Two	I stage: reflexology Vs. meditative practices Vs. standard care II stage: continuing the same therapy Vs. adding the other one	No response on fatigue	(1) Reflexology-Meditation comparison (first stage) (2) Added value of meditation in non-responders to reflexology (3) Added value of reflexology in non-responders to meditation (4) Comparison between the three groups (reflexology-meditation-control group) (5) To identify the optimal intervention sequence by identyifing the optimal decision rule (exploratory objective)	Primary endpoint: Fatigue Secondary endpoints: Summed symptom severity Depression Anxiety	Yes	The study is powered for the second stage. The final sample size is based on the expected response rate of the first stage

Table 1. continued

First author	Number of stages	Interventions per stage	Decision rule	Study objectives	Study endpoints (if reported in the article, distinction is made between primary and secondary endpoint)	Sample size calculation reporting	Does the sample size calculation take into account the multiple stages of randomisation?
Thall et al. [12] (Trials' results)	Two	I stage: randomisation to one of four first-line regimens II stage: randomisation to the remaining second-line regimens	Treatment failure	To identify the best promising regimen to be tested in further confirmatory trials	Overall success/failure	No	
Tummarello et al. [24] (Trials' results)	Two	I stage: CAV-E Vs. CAV-T[a] II stage: rIFN-a-2b Vs. No treatment	Complete response	(1) To compare first-stage treatments (2) To compare second-stage treatments	Tumour response (complete response, partial response, no response) Survival	Yes	No (separate sample size calculation for each stage)
Wang et al. [13] (Trials' results)	Two	I stage: randomisation to one of four first-line regimens II stage: randomisation to the remaining second-line regimens	Treatment failure	To identify the best dynamic treatment regimen to be tested in further confirmatory trials	Primary endpoint: Expert score elicited from the trial's Principal Investigator Secondary endpoints employed for sensitivity analyses: Binary score (overall success/failure) according to the original trial's endpoint Ordinal score	No	

[a]*CAV* cyclophosphamide, doxorubicin and vincristine, *E* etoposide, *T* teniposide.

Table 2. Study analysis information on reports of trial's results and protocols.

First author	Method of analysis	Are the multiple stages taken into account in the analysis?	If multiple stages are taken into account, method specification:	Ancillary analysis
Auyeung et al. [25] (Protocol)	(1) Mean outcome estimation using marginal mean models (2–3) Standard statistical tests	Yes	Marginal mean models. No technical details regarding the estimation are reported	
Bianchi et al. [14] (Trials' results)	Kaplan–Meier method Cox Proportional Hazard model	No. Separate analyses were carried out for the first and second stage		Subgroup analysis
Fisher et al. [15] (Trials' results)	Kaplan–Meier method Log rank test Cox Proportional Hazard Model	No. Separate analyses were carried out for the first and second stage		
Fu et al. [26] (Protocol)	Two-sided Wald-type test Logistic regression Cochran-Armitage trend test Jonckheere-Terpstra trend test	No		
Hammel et al. [16] (Trials' results)	Kaplan–Meier method Cox Proportional Hazard model	No. Separate analyses were carried out for the first and second stage		Sensitivity analysis on time from second randomisation
Hovey et al. [17] (Trials' results)	Kaplan–Meier method Cox Proportional Hazard model	No. Separate analyses were carried out for the first and second stage		
Joss et al. [18] (Trials' results)	Kaplan–Meier method Logistic regression Cox Proportional Hazard model	No. Separate analyses were carried out for the first and second stage		
Kelleher et al. [27] (Protocol)	(1) Two-sample *t*-test ands two-sample test of propotions (2) Approach proposed by Nahu-Shani [34] (3) 3a: Conduction of an overall test testing the null hyporthesis that there are no moderators 3b: if the test rejects the null hypothesis: multiplicity-adjusted statsitical tests using bootrsap methods 3c: Q-learning algorithm and value search estimation (4) Generalised linear models	Yes	Q-learning and value search estimation taking into account of subjects baseline characteristics and response to the first-stage intervention	
Kubota et al. [19] (Trials' results)	Kaplan–Meier method Wilcoxon test Chi-squared test	No		Prognostic factors identification of 3-year survival
Lebeau et al. [20] (Trials' results)	Kaplan–Meier method Log rank test	No. Separate analyses were carried out for the first and second stage. Survival analyses were carried out considering time from second randomisation.		Subgroup analysis
Marshall et al. [21] (Trials' results)	Cumulative incidence competing risk method Kaplan–Meier method Cox Proportional Hazard model	No. Separate analyses were carried out for the first and second stage. venous thromboembolism recurrence analysis was done considering time from second randomisation.		Comparison of patients with no residual deep vein thrombosis and those with residual deep vein thrombosis or pulmonry embolism in the placebo group
Matthay et al. (a) [10] (Trials' results)	Kaplan–Meier method	Yes	The treatment assigned at the second stage was used as grouping variable (transplantation with 13-cis-retinoic acid Vs. transplantation without 13-cis-retinoic acid Vs. chemotherapy with 13-cis-retinoic acid Vs. chemotherapy without 13-cis-retinoic acid). Time from second randomisation was employed in the analysis	Subgroup analysis

Table 2. continued

First author	Method of analysis	Are the multiple stages taken into account in the analysis?	If multiple stages are taken into account, method specification:	Ancillary analysis
Matthay et al. (b) [11] (Trials' results)	Kaplan–Meier method Cox Proportional Hazard model	Yes	The treatment assigned at the second stage was used as grouping variable (transplantation with 13-cis-retinoic acid Vs. transplantation without 13-cis-retinoic acid Vs. chemotherapy with 13-cis-retinoic acid Vs. chemotherapy without 13-cis-retinoic acid). Time from second randomisation was employed in the analysis	Subgroup analysis
Petracci et al. [22] (Trials' results)	Four variants of the Cox model were estimated: (1) two models considering different assumptions on baseline hazards were estimated to assess Bevacizumab efficacy (2) one model including an interaction term between Bevacizumab and rank of disease progression to test the differential effect of Bevacizumab on first and second stage (3) one model including a frailty random effect to test for patients' heterogeneity	Yes	Inverse Probability of Censoring Weighting (IPCW) was estimated to account for potential selection bias at second randomisation. The weights estimation included all covariates that potentially influenced patients' selection. Weights were included in the estimation of the Cox models	Subgroup analysis
Sculier et al. [23] (Trials' results)	Kaplan–Meier method Log rank test Cox Proportional Hazard model	No. Separate analyses were carried out for the second stage		Prognostic factors identification
Sikorskii et al. [28] (Protocol)	Linear mixed effects models or generalised mixed effects models depending on outcomes distribution (obejctives 1–4) Q-learning method (objective 5)	Yes	Q-learning algorithm including two Q functions taking into account dyadic first and second stages history based on regression analyses	
Thall et al. [12] (Trials' results)	Conditional logistic regression Kaplan–Meier method Cox Proportional Hazard model	Yes	Conditional logistic regression for the probability of response in each treatment course, given the previous treatment and outcome history	
Tummarello et al. [24] (Trials' results)	Kaplan–Meier method Log rank test Chi-squared test	No. Separate analyses were carried out for the first and second stage		Subgroup analysis
Wang et al. [13] (Trials' results)	Estimation of the score means for each dynamic treatment regimen tested	Yes	Inverse Probability Treatment Weighting (IPTW) was estimated accounting for patient's history and included in the analysis of the endpoints' means for each dynamic treatment regimen	

of the lack of formal guidelines for designing and reporting trials employing SMART methodologies.

Noticeably, the study design was defined as SMART in only one out of fifteen trial reports included, which would be one of the main reasons why most trials considered each stage separately from the other. This finding could be related to the fact that the formal introduction of SMART design is relatively new, even though the use of multiple randomisations according to pre-specified decision rules dates back to before the 2000s. Consequently, even though such trials, especially the oldest ones, were not labelled as SMART, they presented all the characteristics to be defined as SMART. Except for five, all study protocols and all trials' registrations defined the study design SMART. This difference among the record types could be related to the timing of publication. Trial protocols and trials registrations were published within the last fifteen years, while about one-third of trials' reports were published before 2000.

It is worth pointing out that, despite one of the primary goals of SMART designs is to identify the optimal DTR, only a few records included in the review considered determining the best treatment sequence as the study's primary outcome. Such an aspect detected in the review is reflected by the approaches employed for sizing the study. It is undoubtedly that power analyses for SMART designs present relevant challenges because of the correlation structure between the embedded DTRs [29]. Several approaches have been proposed in the last years to undertake such issues in the SMART design [29–31] without definitive solutions. However, if the primary aim of a SMART study is to identify the best DTR, it follows that the sizing should be done to be able to detect the optimal DTR. However, most of the included trials reporting the power analysis used traditional methods for sample size estimation, since they did not consider the detection of the best DTR as the primary study endpoint. Generally, they estimated the sample size on only one of the trial's stages and inflated the sample size on the expected proportion of subjects entering the second randomisation, or they estimated the sample size for each stage. The present finding is consistent with the conclusions of a recent review in the field [32].

For what concerns data analysis, most of the trials' reports made separate analyses for each trial stage using traditional statistical methods, such as regression-based models, without considering patients' history throughout the study. Such finding is consistent with the fact that most of the records included did not define the study design as SMART and did not identify the evaluation of the optimal DTR among the study outcomes. Focusing on the few reports aimed at identifying the best treatment regimen, two reports of trials' results [13, 22], both reanalyses of previously published data, tried to account for each subjects' history through the trial in treatment effect estimation using propensity weighting estimation, while Thall et al. used a conditional regression approach [12]. On the other side, two study protocols proposed the use of Q-learning algorithm to identify the most promising DTR, which is an approach that has been suggested to be promising for the analysis of data collected using SMART designs [33]. Furthermore, it is interesting to point out that, even though SMART strategies are well known to suffer from the multiple comparison problem since the number of DTRs embedded in the trial is often large, most of the trials included did not account for such a problem.

Finally, it is noteworthy that the focus of SMART designs in solid tumour research changed over time. The first trials published employing sequential multiple assignments were aimed at testing chemo/radio/hormone therapy for solid tumours. Conversely, they have focused more and more on cancer symptoms and cancers treatments side effects, such as fatigue, depression, anxiety and pain in the last years. SMART designs are particularly suitable for assessing complex or long-term interventions for chronic conditions that require management to adapt to patients' needs, as is the case of cancer-related symptoms.

For what concerns study limitations, the search strategy is the main one. When the study was done, no index terms referring to SMART designs were available in any one of the thesaurus of the bibliographic databases searched. Not least, as clearly emerged from the systematic review, the term SMART was often not employed by the authors, even though the study design satisfied the criteria for being SMART. We tried to overcome such limitations by including all possible synonymous of the critical aspects of a SMART design in a well-defined research field, that of solid tumours. However, we cannot rule out that relevant reports could not be included in the search. Another study limitation is that most of the trials' registrations did not report details on the statistical analysis plan. It follows that they contributed to the review only with general information on trial characteristics. However, it would be interesting to update the review to check if reports of these registrations will be published and if the employed methods are consistent with those used in practice.

CONCLUSIONS

The present systematic review showed that the use of SMART designs in solid tumour research is still limited; however, the interest in such designs is growing, and it is testified by the increasing number of protocols and trial registrations in the last years. However, the present work clearly showed that despite the SMART designs' primary aim would be to identify the optimal treatment regimen resulting from the multiple assignments, most of the trials included did not consider the identification of the best DTR as their primary objective. Consequently, they did not employ ad hoc methods for powering and analysing the trial to determine the best DTR; powering and analysing each study's stage separately is still the approach most widely used. Such aspects could be related to the fact that the SMART designs are relatively new.

Present results highlighted that greater efforts should be put forward to developing formal guidelines for SMART designs' conduction and reporting. A thorough literature review of methodological papers presenting and discussing statistical approaches for SMART designs would represent the basis for formal guidelines in the field. With such a review, the development of standard guidelines would benefit from the involvement of a panel of experts, i.e. using the Delphi methodology, to improve the use of such design in cancer trials.

DATA AVAILABILITY

Not applicable

REFERENCES

1. Chakraborty B, Murphy SA. Dynamic treatment regimes. Annu Rev Stat Appl. 2014;1:447–64.
2. Lavori PW, Dawson R. Adaptive treatment strategies in chronic disease. Annu Rev Med. 2008;59:443–53.
3. Kidwell KM. SMART designs in cancer research: past, present, and future. Clin Trials. 2014;11:445–56.
4. Laber EB, Davidian M. Dynamic treatment regimes, past, present, and future: a conversation with experts. Stat Methods Med Res. 2017;26:1605–10.
5. Wahed AS, Thall PF. Evaluating joint effects of induction–salvage treatment regimes on overall survival in acute leukaemia. J R Stat Soc Ser C (Appl Stat). 2013;62:67–83.
6. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Syst Rev. 2021;10:1–11.
7. Kidwell KM. Chapter 2: Dtrs and smarts: Definitions, designs, and applications. In: Adaptive treatment strategies in practice: Planning trials and analyzing data for personalized medicine. SIAM; 2015. p. 7–23.
8. Veritas Health Innovation. Covidence Systematic Review Software. 2021. https://www.covidence.org/.
9. Higgins JP, Savović J, Page MJ, Elbers RG, Sterne JA. Assessing risk of bias in a randomized trial. In: Cochrane handbook for systematic reviews of interventions. The Cochrane Collaboration and John Wiley & Sons Ltd.; 2019, p. 205–28.

10. Matthay KK, Villablanca JG, Seeger RC, Stram DO, Harris RE, Ramsay NK, et al. Treatment of high-risk neuroblastoma with intensive chemotherapy, radiotherapy, autologous bone marrow transplantation, and 13-cis-retinoic acid. N Engl J Med. 1999;341:1165–73.
11. Matthay KK, Reynolds CP, Seeger RC, Shimada H, Adkins ES, Haas-Kogan D, et al. Long-term results for children with high-risk neuroblastoma treated on a randomized trial of myeloablative therapy followed by 13-cis-retinoic acid: a children's oncology group study. J Clin Oncol. 2009;27:1007–13.
12. Thall PF, Logothetis C, Pagliaro LC, Wen S, Brown MA, Williams D, et al. Adaptive therapy for androgen-independent prostate cancer: a randomized selection trial of four regimens. J Natl Cancer Inst. 2007;99:1613–22.
13. Wang L, Rotnitzky A, Lin X, Millikan RE, Thall PF. Evaluation of viable dynamic treatment regimes in a sequentially randomized trial of advanced prostate cancer. J Am Stat Assoc. 2012;107:493–508.
14. Bianchi S, Mosca A, Dalla Volta A, Prati V, Ortega C, Buttigliero C, et al. Maintenance versus discontinuation of androgen deprivation therapy during continuous or intermittent docetaxel administration in castration-resistant prostate cancer patients: a multicentre, randomised Phase III study by the Piemonte Oncology Network. Eur J Cancer. 2021;155:127–35.
15. Fisher B, Dignam J, Bryant J, Wolmark N. Five versus more than five years of tamoxifen for lymph node-negative breast cancer: updated findings from the National Surgical Adjuvant Breast and Bowel Project B-14 randomized trial. J Natl Cancer Inst. 2001;93:684–90.
16. Hammel P, Huguet F, van Laethem JL, Goldstein D, Glimelius B, Artru P, et al. Effect of chemoradiotherapy vs chemotherapy on survival in patients with locally advanced pancreatic cancer controlled after 4 months of gemcitabine with or without erlotinib: The LAP07 Randomized Clinical Trial. JAMA. 2016;315:1844–53.
17. Hovey EJ, Field KM, Rosenthal MA, Barnes EH, Cher L, Nowak AK, et al. Continuing or ceasing bevacizumab beyond progression in recurrent glioblastoma: an exploratory randomized phase II trial. Neuro Oncol Pract. 2017;4:171–81.
18. Joss RA, Alberto P, Bleher EA, Ludwig C, Siegenthaler P, Martinelli G, et al. Combined-modality treatment of small-cell lung cancer: randomized comparison of three induction chemotherapies followed by maintenance chemotherapy with or without radiotherapy to the chest. Ann Oncol. 1994;5:921–8.
19. Kubota K, Furuse K, Kawahara M, Kodama N, Yamamoto M, Ogawara M, et al. Role of radiotherapy in combined modality treatment of locally advanced non-small-cell lung cancer. J Clin Oncol. 1994;12:1547–52.
20. Lebeau B, Chastang C, Allard P, Migueres J, Boita F, Fichet D. Six vs twelve cycles for complete responders to chemotherapy in small cell lung cancer: definitive results of a randomized clinical trial. The "Petites Cellules" Group. Eur Respir J. 1992;5:286–90.
21. Marshall A, Levine M, Hill C, Hale D, Thirlwall J, Wilkie V, et al. Treatment of cancer-associated venous thromboembolism: 12-month outcomes of the placebo versus rivaroxaban randomization of the SELECT-D Trial (SELECT-D: 12m). J Thromb Haemost. 2020;18:905–15.
22. Petracci E, Scarpi E, Passardi A, Biggeri A, Milandri C, Vecchia S, et al. Effectiveness of bevacizumab in first- and second-line treatment for metastatic colorectal cancer: ITACa randomized trial. Ther Adv Med Oncol. 2020;12:1758835920937427.
23. Sculier JP, Paesmans M, Bureau G, Giner V, Lecomte J, Michel J, et al. Randomized trial comparing induction chemotherapy versus induction chemotherapy followed by maintenance chemotherapy in small-cell lung cancer. European Lung Cancer Working Party. J Clin Oncol. 1996;14:2337–44.
24. Tummarello D, Mari D, Graziano F, Isidori P, Cetto G, Pasini F, Santo A, Cellerino R. A randomized, controlled phase III study of cyclophosphamide, doxorubicin, and vincristine with etoposide (CAV-E) or teniposide (CAV-T), followed by recombinant interferon-alpha maintenance therapy or observation, in small cell lung carcinoma patients with complete responses. Cancer. 1997;80:2222–9.
25. Auyeung SF, Long Q, Royster EB, Murthy S, McNutt MD, Lawson D, et al. Sequential multiple-assignment randomized trial design of neurobehavioral treatment for patients with metastatic malignant melanoma undergoing high-dose interferon-alpha therapy. Clin Trials. 2009;6:480–90.
26. Fu SS, Rothman AJ, Vock DM, Lindgren B, Almirall D, Begnaud A, et al. Program for lung cancer screening and tobacco cessation: Study protocol of a sequential, multiple assignment, randomized trial. Contemp Clin Trials. 2017;60:86–95.
27. Kelleher SA, Dorfman CS, Plumb Vilardaga JC, Majestic C, Winger J, Gandhi V, et al. Optimizing delivery of a behavioral pain intervention in cancer patients using a sequential multiple assignment randomized trial SMART. Contemp Clin Trials. 2017;57:51–7.
28. Sikorskii A, Wyatt G, Lehto R, Victorson D, Badger T, Pace T. Using SMART design to improve symptom management among cancer patients: a study protocol. Res Nurs Health. 2017;40:501–11.
29. Artman WJ, Nahum-Shani I, Wu T, Mckay JR, Ertefaie A. Power analysis in a SMART design: sample size estimation for determining the best embedded dynamic treatment regime. Biostatistics. 2020;21:432–48.
30. Almirall D, Lizotte DJ, Murphy SA. SMART Design Issues and the Consideration of Opposing Outcomes: Discussion of "Evaluation of Viable Dynamic Treatment Regimes in a Sequentially Randomized Trial of Advanced Prostate Cancer" by by Wang, Rotnitzky, Lin, Millikan, and Thall. J Am Stat Assoc. 2012;107:509–12.
31. Kim H, Ionides E, Almirall D. A sample size calculator for smart pilot studies. SIAM Undergrad Res Online. 2016;9:229.
32. Bigirumurame T, Uwimpuhwe G, Wason J. Sequential multiple assignment randomized trial studies should report all key components: a systematic review. J Clin Epidemiol. 2022;142:152–60.
33. Almirall D, Nahum-Shani I, Sherwood NE, Murphy SA. Introduction to SMART designs for the development of adaptive interventions: with application to weight loss research. Transl Behav Med. 2014;4:260–74.
34. Nahum-Shani I, Qian M, Almirall D, Pelham WE, Gnagy B, Fabiano GA, et al. Experimental design and primary data analysis methods for comparing adaptive interventions. Psychol Methods. 2012;17:457.

ACKNOWLEDGEMENTS

The work was part of a research project developed in the context of the Master's Programme in Epidemiology of the University of Turin.

AUTHOR CONTRIBUTIONS

GL designed the work, acquired the data, interpreted the results, drafted the work; EP acquired the data, interpreted the results, drafted the work; ES acquired the data and revised the manuscript; IB designed the work, revised the manuscript; DG conceived the work, revised the manuscript; ON conceived the work, revised the manuscript. All authors approved the final version and agreed to be accountable for all aspects of the work.

FUNDING

The authors received no specific funding for this work.

COMPETING INTERESTS

The authors declare no competing interests.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable

CONSENT FOR PUBLICATION

Not applicable

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41416-022-02110-z.

Correspondence and requests for materials should be addressed to Dario Gregori.