Published in final edited form as:

Clin Trials. 2014 August ; 11(4): 445–456. doi:10.1177/1740774514525691.

SMART Designs in Cancer Research: Past, Present and Future

Kelley M. Kidwell

Department of Biostatistics, University of Michigan Ann Arbor, MI 48109 USA

Abstract

Background—Cancer affects millions of people worldwide each year. Patients require sequences of treatment based on their response to previous treatments to combat cancer and fight metastases. Physicians provide treatment based on clinical characteristics, changing over time. Guidelines for these individualized sequences of treatments are known as dynamic treatment regimens (DTRs) where the initial treatment and subsequent modifications depend on the response to previous treatments, disease progression and other patient characteristics or behaviors. To provide evidence-based DTRs, the Sequential Multiple Assignment Randomized Trial (SMART) has emerged over the past few decades.

Purpose—To examine and learn from past SMARTs investigating cancer treatment options, to discuss potential limitations preventing the widespread use of SMARTs in cancer research, and to describe courses of action to increase the implementation of SMARTs and collaboration between statisticians and clinicians.

Conclusions—There have been SMARTs investigating treatment questions in areas of cancer, but the novelty and perceived complexity has limited its use. By building bridges between statisticians and clinicians, clarifying research objectives, and furthering methods work, there should be an increase in SMARTs addressing relevant cancer treatment questions. Within any area of cancer, SMARTs develop DTRs that can guide treatment decisions over the disease history and improve patient outcomes.

1 Introduction

Approximately one in three women and one in two men in the United States will develop cancer in her/his lifetime. Although incidence has declined, cancer remains the leading cause of death worldwide [1]. Simultaneously, advancements in early detection and targeted therapies allow cancer to be caught early and treated aggressively, so more people are survivors of cancer. For all, cancer treatment is ongoing ‘because of the potential persistent and delayed effects of treatment, as well as the risk of recurrence and additional primary malignancies’ [2].

Many cancers are considered chronic diseases because there are effective treatments which prolong progression and metastasis (breast) or because specific cancers naturally develop over a longer period (prostate). Physicians see patients, prescribe a line of therapy and wait for progression. Based on imaging, blood, symptoms or other results, physicians repeatedly make decisions of how best to continue their patients' treatment. This repeated process of observing and treating accounting for disease characteristics and treatment history is common in the treatment of chronic diseases.

Practically, this repeated process is how physicians treat, but mathematically, this is a dynamic treatment regimen (DTR or adaptive treatment strategy). A DTR is a guideline for physicians including sequences of treatments based on intermediate outcomes that are tailored to individuals. DTRs describe treatment of chronic diseases and are applicable to many acute diseases that require tailored sequences of treatment.

An example of a DTR for the treatment of acute myelogenous leukemia (AML) may be: 'After chemotherapy, give a granulocyte-macrophage colony stimulating factor (GM-CSF) and examine bone marrow at 22 days. If there is bone marrow and spinal fluid remission, give cytarabine for four monthly courses. If no remission, no further treatment is given.' A patient will either receive GM-CSF followed by cytarabine or only GM-CSF. Hence, a DTR differs from treatment received by additionally including a treatment path for the patient if he/she had a different intermediate outcome.

Adverse events (toxicities or disease progression, frequent occurrences in oncology trials) should be considered in the construction of DTRs so that the regimens are realistic (Section 2.3). Adding to the DTR above, we specify those with grade III or greater toxicity switch from protocol treatment to a non-specified therapy (or suspend or terminate treatment).

DTRs represent standard clinical practice, basing initial and subsequent treatment on disease-specific prognostic information with subsequent treatment also depending on previous treatment(s). It is not, however, trivial or intuitive to optimize a DTR at the start. Often in treatment of chronic diseases, there are delayed effects [3], where a treatment that produces the best response rate immediately, may end up too toxic for further treatment and result in poor overall response. For example, consider the DTRs in Figure 1 where A looked most promising at first, but (B,C) is the optimal regimen. These delayed effects would be missed in standard one-stage trials [4].

Trials which construct and compare DTRs are known as sequential multiple assignment randomized trials (SMARTs, [4, 5, 6]). Innately, physicians use DTRs in practice; the SMART design allows us to construct and compare DTRs to provide evidence for treatment decisions. A SMART includes two or more stages where patients are randomized to treatment, assessed for an intermediate outcome, and based on this, re-randomized to subsequent treatment. Ideally, a SMART is one in a line of trials that identifies promising DTRs which can then be modified and confirmed in subsequent trials [4]. At the end of the process, an optimal DTR would be used as a treatment guideline for future individuals.

Consider the SMART in Figure 2. This trial enrolled 288 individuals with AML. Patients were randomized to receive GM-CSF or placebo. Patients who achieved remission were randomized to receive cytarabine (C) or cytarabine and mitoxantrone (CM). Patients who did not achieve remission were neither re-randomized nor given further treatment. Within this SMART, there are four embedded DTRs, one of which is mentioned in the DTR example earlier. Note that one DTR contains two treatment paths (DTRs can share paths), but a patient will follow only one path.

For chronic diseases where treatment is ongoing, it is efficient to conduct a SMART where the same individuals receive the entire sequence of treatments to develop and compare DTRs

[8, 9, 3]. SMARTs take advantage of response heterogeneity (investigating subsequent treatment for both responders and non-responder), comorbidities, possibility of relapse, and need for treatment adjustments [10, 11, 12] by design and/or by construction and comparison of DTRs. Specifically, SMARTs can investigate optimal timing of treatments and effects of subsequent treatment in combination with previous treatment and possibilities of stepping down treatment due to poor adherence, side-effects or other burdens. SMARTs can address a wider variety of treatment questions as opposed to the standard randomized controlled trial (RCT) while more closely mimicking the treatment process.

Decisions must be made at each stage as to whether re-randomization is appropriate, but repeated randomization is justified by questions of optimal treatment in sequence. Randomization serves to answer research questions and justify causal inferences within a SMART, but is not included in the resulting formulation of DTRs. Randomization can be based on baseline characteristics and response to previous treatment abiding by standard ethics of trials [13]. It is advantageous to re-randomize at each stage using information up to randomization for balance, but upfront randomization to DTRs is possible [14].

Data from a SMART includes patient and disease characteristics at baseline and throughout treatment, treatment received at all stages, and intermediate outcomes. This data can be utilized to construct DTRs by employing methods from causal inference and using the potential outcomes or counterfactual framework [15, 16]. Potential outcomes are connected to the observed SMART data under consistency and no unmeasured confounders assumptions [17] to evaluate and estimate optimal DTRs. This framework allows comparison of the distributions of survival times if all patients were assigned to DTRs of interest and addresses intention-to-treat questions (refer to [18, 6, 4, 19, 20]).

The SMART design exploits the fact that combinations and sequences of therapies are likely the best course of treatment, but individual treatments may have different effects when used in combination or sequence with others. This is contingent on investigating response to treatment which occurs in a timely fashion so that randomizations occur within a relatively short interval. For many cancers, this is not an issue, however, it may limit the value of SMARTs for some types of disease. In cases with long natural histories where sequential treatment is beneficial, it is important to develop methods for valid inference from observational data [21, 22, 23]. Here, however, we focus on cancers which can be studied sequentially in a timely manner.

In Section 2 we learn from cancer SMARTs and present current limitations to the expanded use of SMARTs in Section 3. Aims are discussed in Section 4 to advance the understanding, implementation, methodology and collaboration for increased use of SMARTs. A brief discussion concludes promoting SMARTs and development of DTRs broadly, but principally in cancer research.

2 Learning from the Past

Many precursors to SMARTs were not specifically designed as SMARTs or to analyze DTRs. Rather, these trials were designed to answer inherent questions requiring two stages with the second-stage treatment depending on the first-stage response.

SMART designs have been implemented in small-cell lung [24, 25], AML [7, 26, 27], neuroblastoma [28, 29], lymphoma [30, 31, 32], prostate [33], multiple myeloma [34], malignant melanoma [35] and renal cell carcinoma [36]. Each of these trials contained two stages with initial randomization to a treatment (2-4 treatments) and follow-up to assess initial response (3 weeks to 8.5 months). Based on this response (no progressive disease, complete response, remission, clinical definition of therapy success), all or a subset were randomized to a second-stage treatment (switch to a different treatment, adding therapy, continuing or discontinuing previous treatment, or no further treatment). Using these trials as examples, we extend beyond their aims to recognize appropriate questions within the DTR framework and compare results.

2.1 SMARTs can investigate sequential and stage-specific treatment effects

First we must determine DTRs used in practice and DTRs which may not be currently used, but are promising. Once identified, we embed these in a SMART. Then we can make comparisons between tailored sequences of treatment, not just compare treatments at a particular time for a subset of individuals. Clinically, this translates to evidence-based development of a transcript of treatment decisions that physicians refer to in seeking the best overall outcome for patients. A SMART can address stage-specific, as well as, sequential, DTR-based treatment effects. Table 2.1 enumerates the possible aims of a SMART [11].

For patients, overall survival is the most important endpoint, not the value of the individual steps. Traditionally, we assume that the whole is equal to the sum of the parts, but this is not always true. Stage-specific or local comparisons may be interesting and useful, but they can also be myopic. Stringing together the best treatments at each stage does not always lead to the best overall treatment strategy. Antagonistic or synergistic treatment effects occur, but cannot be examined except in the setting of a SMART where the same individuals are followed.

Consider the trial investigating the treatment of children with high-risk neuroblastoma from Matthay et al. [28] which includes four embedded DTRs. In the 1999 article, the authors performed a test of differences in survival at three years in marginal groups (conditional on no disease progression) who followed one of four treatment paths. We present survival curves where the left figure is from the original text and the right figure is the weighted version for those following each DTR. The DTR-based figure differs from the original by including those who progressed and measures time after first randomization. There is a notable difference in the number at risk since DTRs take advantage of all those enrolled. Specifically, numbers given in Figure 3 are subsets of patients who followed the given treatment path (for example, 50 patients had transplant, no disease progression and received 13-cis-retinoic acid), whereas numbers in Figure 4 include all who were consistent with a given DTR (along with 50 patients who received 13-cis-retinoic acid, 90 had disease

progression). In Figure 3 each treatment path is mutually exclusive such that patients can be at risk in one group, but in Figure 4, patients can be consistent with multiple DTRs so their information is utilized in estimation by inverse-probability-of-treatment-weighting (IPTW). Notice the separation between curves after two years from the second randomization in the original figure is much less apparent in the DTR-based figure.

The 1999 paper made two local comparison conclusions: (1) bone transplantation is advantageous over chemotherapy, $p=0.034$; (2) 13-cis-retinoic acid is advantageous over no further treatment for those without disease progression regardless of induction therapy, $p=0.027$. We find discrepancy between local and global comparisons, however, as an omnibus weighted log-rank test concludes that there is no significant difference in survival between any of the DTRs, $p=0.57$ [47] (mirrored by overlapping curves in Figure 4).

In this DTR-based method those who follow the regimen of interest are re-weighted using IPTW to account for the loss of information due to some patients following other regimens as in a missing data problem. This creates a sample representative of all patients following DTRs of interest. These weights account for randomization probabilities, response status, and censoring [18]. IPTW can additionally help to achieve balance, even in the randomized setting. The original analysis, in contrast, included unweighted subsets of patients following particular treatment paths. For further analysis of this data with similar DTR-based conclusions, see the following using associated weighted methods: Cox regression [42], Kaplan-Meier estimates [48], risk set estimator [38], and marginal mean estimators [18].

Even when using a sequential two-stage design, we may not be able to piece together stage-specific results to mimic DTRs used in practice. We may shed light on DTRs of interest, but constructing and comparing DTRs within a SMART utilizing methods developed in this area provide the best evidence for clinical application of personalized sequential treatment. By conducting SMARTs we avoid faulty conclusions, thereby strengthening translation from trials to clinic.

2.2 SMARTs allow for efficient use of individuals

Consider a trial for small cell lung carcinoma [25] in Figure 5 to illustrate the difference in the number of patients used in a marginal versus DTR analysis. The authors could not make conclusions about maintenance treatment due to the small number who experienced complete response (this emphasizes the usefulness of pilot studies, Section 4). The exact numbers who received maintenance treatments were not available, but the advantage of analyzing DTRs as opposed to stage-specifically is apparent. Instead of conditioning on response and having very small groups, survival could be estimated for DTRs, which have increased size. For the DTR CAVE-E-RIFA, individuals who received RIFA after CAV-E and individuals who did not respond to CAV-E are combined (at least 60 individuals). Using IPTW methods, information is used from those who did not achieve a complete response (weight=1) and those who did (weight is inverse of probability of getting second-stage treatment of interest), boosting sample size and allowing for efficient DTR conclusions. Note that this gain in effective sample size from the (non-)responders must be matched by a reduction in the total number of strategies compared [14].

2.3 Definition of DTR must be explicit

DTRs must be defined in terms of their clinical utility and can depend on efficacy, toxicity and disease progression which are additional outcomes in trials. To take advantage of all who participate in a SMART, DTRs must also include what happens to those who respond but have high toxicities and therefore cannot receive subsequent treatment. To illustrate this consider a prostate cancer SMART [49] which included four treatments (CVD, KA/VE, TEE, TEC). The treatment algorithm includes 12 possible two-stage strategies (see [49]).

The original stage-specific analysis concluded that TEC was the best initial and KA/VE the best second treatment [49]. Data re-analyzed using IPTW found success rates were highest for the DTRs (CVD,KA/VE), (TEC,CVD), and (TEE,CVD) [50]. Both of these analyses, however, assumed non-informative drop-out, however 47/150 (31.3%) participants did not complete their therapy per algorithm, 35 for toxicity and/or progressive disease. This led to redefining DTRs to include a switch to non-specified therapeutic/palliative care for those who developed toxicity or progressive disease [43]. These redefined or viable DTRs elucidate the need for considering and appropriately analyzing DTRs which are likely to include drop-out. Embedded DTRs should be declared viable prior to the conduct of SMARTs preparing for possible participant issues. A pilot study will eliminate the burden of complexity that these definitions may invoke in planning (Section 4).

In summary, SMARTs can address stage-specific questions, but can also address sequential treatment effects highlighting the advantages of developing viable DTRs for clinical use. More recent SMARTs, [36] and [35], were designed to develop and compare DTRs. These designs, analysis plans, and eventual results provide concrete examples of cancer SMARTs addressing global treatment questions by developing personalized decision rules. Current and past trials in cancer provide valuable information, but we can also gain insight from SMARTs conducted in other areas (<http://methodology.psu.edu/ra/adap-treat-strat/projects>).

3 Current Limitations

There have only been a small number of cancer SMARTs in large part due to the novelty of this area and short time period which has passed for development and dissemination of methods. Due to new vocabulary and intricate-looking designs, SMARTs and their embedded DTRs can be intimidating. There are three main perceived limitations that inhibit the expanded use of SMARTs.

3.1 Physicians' perception of complexity and large sample size

Physicians rooted in cancer research are familiar with the standard RCT. Thall explains, 'Tailoring a design to fit particular trial is time-consuming... the investigators must sell the design to colleagues in their own clinics, to physicians in other medical centers considering participating in a multi-institution study, and to regulatory agencies. Given these difficulties, it is not surprising that many investigators choose a conventional design... The motivation for an investigator to choose a more complicated but more realistic trial design comes from the belief that there is something substantive to be gained by its use, and the investigator's trust in his/her chosen statistician's abilities [51].' Given the more clinically useful

conclusions about personalized treatment plans from SMARTs, we must actively motivate physicians and statisticians to work together to pursue this design.

We predominantly think of treatment or group comparison as the objective of clinical trials, but that does not need to be the case. Feasibility, development, estimation and personalization may instead or additionally be a trial's goal(s). Generally, new designs are met with hesitation, but in other areas they have successfully flourished, especially with increased attention to genomics, biomarkers, and adaptive design. Perceived complexity and apparent large sample size of SMARTs may just be an illusion which can be clarified by establishing the trial's primary objective(s).

In designing a SMART, like any other clinical trial, the primary and secondary objectives must be well defined. Objectives of a SMART may be exactly the same as an RCT, or exceed those by constructing, comparing or optimizing DTRs (Table 2.1). If stage-specific comparisons are the primary aims, the trial is powered similarly to a RCT, but as a bonus, DTRs can be explored. If the goal is based on embedded DTRs, sample size must account for the multiple randomizations. The total number needed for a SMART is similar to that from multiple single-stage RCTs, but that number is needed upfront. Wolbers and Helderbrand [52] discuss this issue and present simulation results comparing three single-stage RCTs versus one SMART. They conclude if the primary aim is centered around the initial treatment and there is low probability of maintenance success, then a one-stage design is faster and more efficient. If both initial and maintenance treatment performance is in question, a SMART is more efficient by utilizing all individuals and generally takes less time with similar total sample size. Others echo these conclusions, supporting the increased use of SMARTs [14] with focus on cost [53, 54]. Sample size methods for survival analysis of embedded DTRs are presented in [41], [55], [56], while methods for continuous or binary outcomes can be found in [57], [58], [59], [4], and [11]. If the primary aim is to tailor DTRs by baseline or time-varying characteristics, then Q-learning [45, 60, 46] is required which has no published sample size methods as of now.

SMARTs just like any trial can be complex, large and intimidating, but clarification of aims simplifies and guides the process. Any successful trial demands a close relationship between statisticians, physicians and other researchers, so this should not dissuade the use of SMARTs.

3.2 Pharmaceutical companies' dislike of drug comparison

Beyond the reluctance to launch a novel trial design, institutions are hesitant to approach pharmaceutical companies to fund research which may combine or compare drugs in sequences with other companies' drugs. Although it is possible to write in funding in NIH-funded grants, pharmaceutical backing is virtually essential in any cancer drug trial. Pharmaceutical support may seem problematic, but there have been cases in mental health where there was significant support of cross-company drug comparisons (STAR*D [61] and CATIE [62]) due to the breadth, scope, and innovation of these trials.

Pharmaceutical companies may be more likely to partner with academic institutions in pursuit of implementing SMARTs if their drugs are used in: combination or compared to the

standard of care; the maintenance stage when primary use was established as front-line therapy; off-label in an area which was not its primary approval; a large trial likely to change and improve treatment. As drug development slows and requires extensive funding, SMARTs provide options to find new uses for existing drugs. These issues are not new nor unique to SMARTs, but with more treatment branches, the more likely it is to combine, compare and find new applications for existing treatments.

To motivate pharmaceutical support, it may be as simple as introducing this area to companies explaining the benefits of SMARTs in the treatment of individuals, the prescription process for physicians, and the profit for companies by providing evidence for treatments in sequences. This may be done by presenting the following scenario to company with drug A where interest is in two-stage DTRs sequencing A with drugs B and C from other companies for non-responders. DTRs of interest include $\{(A,B),(A,C),(B,A),(B,C),(C,A),(C,B)\}$. A SMART interested in these six DTRs uses drug A in 67% of the regimens which is likely to exceed half of the patients. Thus, not only could more patients receive drug A than in a two or three-arm one-stage trial, but if the drug is expected to be used in sequence, a SMART would address relevant clinical questions.

3.3 Lack of methods

Lastly, a significant impediment to the increased use of SMARTs is the lack of statistical methods. There are numerous methods, but more statistical theory is necessary. There are open questions in design and analysis for easy to implement sample size calculators, different types of outcomes, longitudinal and joint analyses, further applications of Q-learning, and understanding how to define and include clinically meaningful tailoring variables. This presents the perfect opportunity for physicians and statisticians to join forces for clinical and statistical grants in DTR development.

4 Looking to the Future

To encourage implementation and diminish perceived limitations of SMARTs, we must collaborate and then disseminate this information. Looking toward the future, we hope to improve individuals' lives by tailoring treatment throughout diagnosis, survivorship and relapse. SMARTs can play an integral part if we build bridges across disciplines, foster implementation at the pilot study level, continue methods research, increase the efficiency of existing drugs by using them in new ways, and encourage collaboration.

Statisticians have done a tremendous job creating and cultivating this area, but we must occasionally step away from our desks to introduce and advocate. Our methods are only as good as their relevance, understand-ability and accessibility to others from all backgrounds. Thus, manuscripts which are approachable from the viewpoint of physicians, nurses, research assistants, patient advocates and executives are necessary. We need to build bridges from statistical methods to clinical applications. Pieces such as [50], [10], [6], [63], [12] are imperative to bridge the gap between methods to their use and consumption. Susan Murphy's group has disseminated methods by providing relevant, non-technical articles, personal websites listing research and presentations, an informative group website (<http://methodology.psu.edu/ra/adap-inter>), workshops, and brainstorming sessions for those

interested in implementing SMART designs in behavioral health. We must follow suit and build bridges in cancer.

Manuscripts in clinical journals are a critical step. Simulations or previous trials may be used to illustrate SMART properties and advantages. Additionally, posters and conversations with clinicians/researchers internally or at conferences are indispensable. We cannot be surprised by the lack of cancer SMARTs if we do not advocate their practicality and enlighten others to their accessibility.

We must get out from behind our desks and advocate about SMART design, but then we must also sit back down to further methods research or we must value those who do both. Through collaboration, areas with gaps will become obvious and provide for timely and comprehensible methods. This area allows statisticians to show their great abilities to be both methodologists and collaborators. It is important to bring this area to the attention of statisticians, other researchers and clinicians at every venue. The more collaboration, the more innovative we can be in addressing statistical theory.

In the process of building bridges, pilot studies play a pivotal role and ‘can enhance the likelihood of success of the main study and potentially help to avoid doomed main studies’ [64]. Physicians and statisticians can obtain experience conducting a SMART while assessing its feasibility and logistics. Pilot studies are the gateway to experience and gaining understanding about new designs. The NIH supports pilot studies through R21, R34 and K mechanisms which can lead to R01 proposals involving SMARTs. An outstanding resource for designing a SMART pilot study is available by Almirall et al. [65].

Finally, creating a **SMARTer Cancer Care Consortium** could lead to incredible statistical and clinical advancements. This could be a standalone consortia defined by involved universities or a group of individuals from several universities that collaborate with existing consortia to design and implement SMARTs while simultaneously advancing methods. There would be heavy statistical influence, but the group would be composed of all types of researchers. By uniting academic institutions and hospital systems (and perhaps pharmaceutical companies), an increased number of larger and more innovative personalized cancer trials could result. Such an entity will most certainly bring attention to this area while increasing experience, methods and cancer treatment. This type of group, or at least cross-university collaboration and multi-site SMARTs should be a priority to foster patient care.

5 Discussion

The SMART design is a natural choice for investigating cancer treatment questions because most patients require sequential treatments. SMARTs address the timing, sequencing and tailoring of treatments and provide guidelines to inform clinical judgment. SMARTs efficiently use information from all individuals following DTRs, randomizing the same individuals throughout, and may reach clinical decisions in less time compared to multiple single-stage trials while answering more complex and relevant treatment questions. By establishing the primary aims of a SMART, clarity, efficiency, and unbiased results follow with a comparable sample size to standard trials.

We have learned from past SMARTs and are now more equipped to analyze resulting DTR data. Statisticians and clinicians need to work closely together to continue to develop and promote this area. With increased implementation of the SMART design, physicians, statisticians and pharmaceutical companies will become more familiar with SMART benefits. By encouraging collaboration and presenting accessibility, the SMART will become a routine trial providing innovative and personalized treatment strategies to improve cancer patient outcomes.

Acknowledgments

Thank you to Abdus S. Wahed, Jane Perlmutter, Jonas Ellenberg, and the two reviewers for their incredibly helpful feedback.

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

References

1. Globocan. Globocan 2008 Section of Cancer Information. 2013. <http://www.who.int/mediacentre/factsheets/fs297/en/index.html>
2. Siegel R, DeSantis C, Virgo K, Stein K, Mariotto A, Smith T, et al. Cancer treatment and survivorship statistics, 2012. CA: A Cancer Journal for Clinicians. 2012; 62(4):220–241. Available from: <http://dx.doi.org/10.3322/caac.21149>. [PubMed: 22700443]
3. Chakraborty B. Dynamic Treatment Regimes for Managing Chronic Health Conditions: A Statistical Perspective. American Journal of Public Health. 2011; 101:40–45. [PubMed: 21088260]
4. Murphy SA. An experimental design for the development of adaptive treatment strategies. Statistical Methods. 2005; 24:1455–81.
5. Lavori PW, Dawson R, J RA. Flexible treatment strategies in chronic disease: Clinical research implications. Biological Psychology. 2000; 48:605–614.
6. Lavori PW, Dawson R, J RA. Dynamic treatment regimes: Practical design considerations. Clinical Trials. 2004; 1:9–20. [PubMed: 16281458]
7. Stone RM, Berg DT, George SL, Dodge RK, Paciucci PA, Schulman PP, et al. Granulocyte-macrophage colony-stimulating factor after initial chemotherapy for elderly patients with primary acute myelogenous leukemia. New England Journal of Medicine. 1995; 332:1671–1677. [PubMed: 7760868]
8. Lavori PW, Dawson R. Adaptive treatment strategies in chronic disease. Annual Review of Medicine. 2008; 59:443–453.
9. Murphy SA, Oslin D, Rush SJ, Zhu J. Methodological Challenges in Constructing Effective Treatment Sequences for Chronic Psychiatric Disorders. Neuropsychopharmacology. 2007; 32:257–262. [PubMed: 17091129]
10. Lei H, Nahum-Shani I, Lynch K, Oslin D, Murphy SA. A “SMART” Design for building individualized treatment sequences. Annual Review of Clinical Psychology. 2012; 8:21–48.
11. Oetting, AI.; Levy, JA.; Weiss, RD.; Murphy, SA. Statistical Methodology for a SMART Design in the Development of Adaptive Treatment Strategies. In: Shrout, P.; Keyes, K.; Ornstein, K., editors. Causality and Psychopathology: Finding the Determinants of Disorders and their Cures American Psychopathological Association. Arlington, VA: American Psychiatric Publishing, Inc.; 2011. p. 179–205.
12. 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. Psychological Methods. 2012; 17:457–77. [PubMed: 23025433]
13. Rosner F. The ethics of randomized clinical trials. The American Journal of Medicine. 1987; 82:283–290. [PubMed: 3812521]

14. Ko JH, Wahed AS. Up-front versus sequential randomizations for inference on adaptive treatment strategies. *Statistics in Medicine*. 2012; 31:812–830. [PubMed: 22362642]
15. Robins JM. A new approach to causal inference in mortality studies with sustained exposure periods- application to control of the healthy worker survivor effect. *Mathematical Modelling*. 1986; 7:1393–1512.
16. Rubin DB. Estimating causal effects of treatments in randomized and non-randomized studies. *Journal of Educational Psychology*. 1974; 66:688–701.
17. Robins, J. Causal Inference from Complex Longitudinal Data. In: Berkane, M., editor. *Latent Variable Modeling and Applications to Causality* vol 120 of *Lecture Notes in Statistics*. Springer; New York: 1997. p. 69-117. Available from: http://dx.doi.org/10.1007/978-1-4612-1842-5_4
18. Lunceford JK, Davidian M, Tsiatis AA. Estimation of survival distributions of treatment policies in two-stage randomization designs in clinical trials. *Biometrics*. 2002; 58:48–57. [PubMed: 11890326]
19. Bembom O, van der Laan MJ. Analyzing Sequentially Randomized Trials based on causal effect models for realistic individualized treatment rules. *Statistics in Medicine*. 2008; 27:3689–3716. [PubMed: 18407580]
20. Chakraborty B, Murphy SA. Dynamic Treatment Regimes. *Annual Review of Statistics and its Applications*. 2014; 1:1.1–1.18.
21. Huang X, Cormier JN, Pisters PWT. Estimation of the Causal Effects on Survival of Two-Stage Nonrandomized Treatment Sequences for Recurrent Diseases. *Biometrics*. 2006; 62:901–909. [PubMed: 16984334]
22. Hernn MA, Lanoy E, Costagliola D, Robins JM. Comparison of Dynamic Treatment Regimes via Inverse Probability Weighting. *Basic & Clinical Pharmacology & Toxicology*. 2006; 98(3):237–242. Available from: http://dx.doi.org/10.1111/j.1742-7843.2006.pto_329.x. [PubMed: 16611197]
23. Li, Z.; Valenstein, M.; Pfeiffer, P.; Ganoczy, D. A global logrank test for adaptive treatment strategies based on observational studies; *Statistics in Medicine*. 2013. p. n/a-n/a. Available from: <http://dx.doi.org/10.1002/sim.5987>
24. 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. *Annals of Oncology*. 1994; 5:921–928. [PubMed: 7696164]
25. Tummarello D, Mari D, Graziano F, Isidori P, Cetto G, Pasini F, et al. 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; 80:2222–2229. [PubMed: 9404698]
26. Stone RM, Berg DT, George SL, Dodge RK, Paciucci PA, Schulman PP, et al. Postremission therapy in older patients with de novo acute myeloid leukemia: A randomized trial comparing mitoxantrone and intermediate-dose cytarabine with standard-dose cytarabine. *Blood*. 2001; 98:548–53. [PubMed: 11468148]
27. Thomas X, Raffoux E, Botton Sd, Pautas C, Arnaud P, de Revel T, et al. Effect of priming with granulocytemacrophage colony-stimulating factor in younger adults with newly diagnosed acute myeloid leukemia: a trial by the Acute Leukemia French Association (ALFA) Group. *Leukemia*. 2007; 21:453–461. [PubMed: 17252021]
28. 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. *The New England Journal of Medicine*. 1999; 341:1165–1173. [PubMed: 10519894]
29. 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. *Journal of Clinical Oncology*. 2009; 27:1007–1013. [PubMed: 19171716]

30. Habermann TM, Weller EA, Morrison VA, Gascoyne RD, Cassileth PA, Cohn JB, et al. Rituximab-CHOP versus CHOP alone or with maintenance Rituximab in older patients with diffuse large B-cell lymphoma. *Journal of Clinical Oncology*. 2006; 24:3121–3127. [PubMed: 16754935]
31. van Oers MHJ, Klasa R, Marcus RE, Wolf M, Kimby E, Gascoyne RD, et al. Rituximab maintenance improves clinical outcome of relapsed/resistant follicular non-Hodgkin lymphoma in patients both with and without rituximab during induction: results of a prospective randomized phase 3 intergroup trial. *Blood*. 2006; 108:3295–301. [PubMed: 16873669]
32. Forstpointner R, Unterhalt M, Dreyling M, Bock HP, Repp R, Wandt H, et al. Maintenance therapy with rituximab leads to a significant prolongation of response duration after salvage therapy with a combination of rituximab, fludarabine, cyclophosphamide, and mitoxantrone (R-FCM) in patients with recurring and refractory follicular and mantle cell lymphomas: results of a prospective randomized study of the German Low Grade Lymphoma Study Group (GLSG). *Blood*. 2006; 108:4003–4008. [PubMed: 16946304]
33. Thall PF, Millikan RE, Sung HG. Evaluating multiple treatment courses in clinical trials. *Statistics in Medicine*. 2000; 30:1011–1128.
34. Mateos MV, Orio A, Martínez-López J, Gutiérrez N, Teruel AI, de Paz R, et al. Bortezomib, melphalan, and prednisone versus bortezomib, thalidomide, and prednisone as induction therapy followed by maintenance treatment with bortezomib and thalidomide versus bortezomib and prednisone in elderly patients with untreated multiple myeloma: a randomised trial. *The Lancet Oncology*. 2010 Oct.11:934–941. [PubMed: 20739218]
35. 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. *Clinical Trials*. 2009; 6:480–490. [PubMed: 19786415]
36. Thall PF, Wooten LH, Logothetis CJ, Millikan RE, Tannir NM. Bayesian and frequentist two-stage treatment strategies based on sequential failure times subject to interval censoring. *Statistics in Medicine*. 2007; 26:4687–4702. [PubMed: 17427204]
37. Wahed AS, Tsiatis AA. Optimal estimator for the survival distribution and related quantities for treatment policies in two-stage randomization designs in clinical trials. *Biometrics*. 2004; 60:124–133. [PubMed: 15032782]
38. Guo X, Tsiatis AA. A weighted risk set estimator for survival distributions in two-stage randomization designs with censored survival data. *The International Journal of Biostatistics*. 2005; 1:1–15.
39. Wahed AS, Tsiatis AA. Semiparametric efficient estimation of survival distributions in two-stage randomisation designs in clinical trials with censored data. *Biometrika*. 2006; 93:163–177.
40. Lokhnygina Y, Helderbrand JD. Cox regression methods for two-stage randomization designs. *Biometrics*. 2007; 63:422–428. [PubMed: 17425633]
41. Feng W, Wahed AS. Supremum weighted log-rank test and sample size for comparing two-stage adaptive treatment strategies. *Biometrika*. 2008; 95:695–707.
42. Tang X, Wahed AS. Comparison of treatment regimes with adjustment for auxiliary variables. *Journal of Applied Statistics*. 2011; 38:2925–2938.
43. 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. *Journal of the American Statistical Association*. 2012; 107
44. Wahed AS, Thall PF. Evaluating joint effects of induction/salvage treatment regimes on overall survival in acute leukaemia. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*. 2013; 62:67–83.
45. Zhao Y, Zeng D, Socinski MA, Kosorok R Michael. Reinforcement Learning Strategies for Clinical Trials in Nonsmall Cell Lung Cancer. *Biometrics*. 2011; 67:1422–1433. [PubMed: 21385164]
46. Goldberg Y, Kosorok MR. Q-learning with censored data. *The Annals of Statistics*. 2012; 40:529–560. [PubMed: 22754029]
47. Kidwell KM, Wahed AS. Weighted Log-Rank Statistic to Compare Shared-Path Adaptive Treatment Strategies. *Biostatistics*. 2013; 14(2)

48. Miyahara S, Wahed AS. Weighted Kaplan-Meier estimators for two-stage treatment regimes. *Statistics in Medicine*. 2010; 29:25812591.
49. 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. *Journal of the National Cancer Institute*. 2007; 99(21):1613–1622. [PubMed: 17971530]
50. Bembom O, van der Laan MJ. Statistical Methods for Analyzing Sequentially Randomized Trials. *Journal of the National Cancer Institute*. 2007; 99:1577–82. [PubMed: 17971533]
51. Thall PF. A review of phase 2-3 clinical trial designs. *Lifetime Data Analysis*. 2008; 14:37–53. [PubMed: 17763973]
52. Wolbers M, Helderbrand JD. Two-stage Randomization Designs in Drug Development. *Statistics in Medicine*. 2008; 27:4161–4174. [PubMed: 18570274]
53. Buyze J, Rompaye BV, Goetghebeur E. Designing a sequentially randomized study with adherence enhancing interventions for diabetes patients. *Statistics in Medicine*. 2010; 29:1114–1126. [PubMed: 20101597]
54. Buyze J, Goetghebeur E. Evaluating dynamic treatment strategies: does it have to be more costly? *Pharmaceutical Statistics*. 2012; 12:35–42. [PubMed: 23208788]
55. Feng W, Wahed AS. Sample size for two-stage studies with maintenance therapy. *Statistics in Medicine*. 2009; 28:2028–2041. [PubMed: 19382105]
56. Li Z, Murphy SA. Sample size formulae for two-stage random trials with survival outcomes. *Biometrika*. 2011; 98:503–518. [PubMed: 22363091]
57. Orellana L, Rotnitzky A, Robins JM. Dynamic Regime Marginal Structural Mean Models for Estimation of Optimal Dynamic Treatment Regimes, Part I: Main Content. *The International Journal of Biostatistics*. 2010; 6
58. Dawson R, Lavori PW. Placebo-free designs for evaluating new mental health treatments: The use of adaptive treatment strategies. *Statistics in Medicine*. 2004; 23:3249–3262. [PubMed: 15490427]
59. Dawson R, Lavori PW. Sample Size Calculations for Evaluating Treatment Policies in Multi-Stage Designs. *Clinical Trials*. 2010; 7:643–652. [PubMed: 20630903]
60. Zhao Y, Zeng D, Socinski MA, Kosorok R Michael. Estimating Individualized Treatment Rules Using Outcome Weighted Learning. *Journal of the American Statistical Association*. 2012; 107:1106–1118. [PubMed: 23630406]
61. Rush AJ, Fava M, Wisniewski SR, Lavori PW, Trivedi MH, Sackeim HA, et al. Sequenced Treatment Alternatives to Relieve Depression (STAR*D): Rationale and design. *Controlled Clinical Trials*. 2004; 25:119–42. [PubMed: 15061154]
62. Schneider LS, Tariot PN, Lyketsos CG, Dagerman KS, Davis KL, Davis S, et al. National Institute of Mental Health Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE): Alzheimer disease trial methodology. *The American Journal of Geriatric Psychiatry*. 2001; 9:346–360. [PubMed: 11739062]
63. Murphy SA, Lynch KG, McKary JR, Oslin D, Tenhave T. Developing Adaptive Treatment Strategies in Substance Abuse Research. *Drug and Alcohol Dependence*. 2007; 88:S24–S30. [PubMed: 17056207]
64. Thabane L, Ma J, Chu R, Cheng J, Ismaila A, Rios LP, et al. A tutorial on pilot studies: the what, why and how. 2010; doi: 10.1186/1471-2288-10-1
65. Almirall D, Compton SN, Gunlicks-Stoessel M, Duan N, Murphy SA. Designing a pilot sequential multiple assignment randomized trial for developing an adaptive treatment strategy. *Statistics in Medicine*. 2012; 31:1887–1902. [PubMed: 22438190]

Figure 1. Two dynamic treatment regimens, (A,C) and (B,C) where the initial response rate is best for treatment A, but the overall best response rate is for DTR (B,C).

Figure 2.
SMART design from an acute myelogenous leukemia trial [7] including 4 embedded DTRs.

No. AT RISK							
Transplantation, 13- <i>cis</i> -retinoic acid	50	33	25	13	10	4	1
Transplantation, no 13- <i>cis</i> -retinoic acid	48	28	18	11	6	5	1
Chemotherapy, 13- <i>cis</i> -retinoic acid	52	21	16	12	9	6	3
Chemotherapy, no 13- <i>cis</i> -retinoic acid	53	26	13	7	6	4	2

Figure 3.
Survival curves of marginal treatment regimens from [28].

Figure 4.
Survival curves of four DTRs using weighted risk set estimator.

Figure 5. SMART design from a small cell lung carcinoma trial [25]. There are four embedded DTRs in this SMART which will be denoted by CAV-E-RIFA, CAV-T-RIFA, CAV-E-N, CAV-T-N. The DTR CAV-E-RIFA reads as, ‘treat with CAV-E followed by recombinant interferon-alpha (RIFA) if complete response, continue CAV-E if partial or no response’. Another DTR, CAV-E-N, reads as, ‘treat with CAV-E followed by nothing (observation) for those with complete response and continue CAV-E for those without complete response.’

Table 1

Possible Aims of a SMART. Aims (1) and (2) are similar to questions addressed by standard one-stage randomized clinical trials (RCTs), but aims (3) and (4) give further insight into dynamic treatment regimens. All are possible through a sequential multiple assignment randomized trial (SMART). References provided are to methods and applied work for survival outcomes.

Aim or Clinical Question	RCT	SMART	References
1. What is the best first-stage treatment (this may be a comparison of two or more treatments or of different time schedules, doses or delivery methods)?	X	X	usual RCT methods
2. What is the best second-stage treatment for (non-)responders to a particular first-stage treatment?	X	X	usual RCT methods
3. Does the outcome significantly differ between two or more DTRs (this may be a hypothesis test or estimation)?		X	[18, 37, 38, 39, 40, 41, 42, 43, 44]
4. For a specific DTR, can we improve individual outcomes by further tailoring treatment by baseline or time-varying characteristics?		X	[45, 46]