DIA DEVELOP INNOVATE ADVANCE

Therapeutic Innovation & Regulatory Science
2017, Vol. 51(1) 77-88

DOI: 10.1177/2168479016665434
tirs.sagepub.com

DIA's Adaptive Design Scientific Working Group (ADSWG): Best Practices Case Studies for "Less Well-understood" Adaptive Designs

Eva Miller[1], Paul Gallo[2], Weili He[3], Lisa A. Kammerman[4], Kenneth Koury[3], Jeff Maca[5], Qi Jiang[6], Marc K. Walton[7], Cunshan Wang[8], Katherine Woo[7], Cynthia Fuller[9], and Yannis Jemiai[10]

Abstract

Adaptive design (AD) clinical trials use accumulating subject data to modify the parameters of the design of an ongoing study, without compromising the validity and integrity of the study. The 2010 US Food and Drug Administration (FDA) Draft Guidance on Adaptive Design Clinical Trials described a subset of 7 primary design types as "less well-understood." FDA defined these designs as those with limited regulatory experience. To better understand the properties of these less well-understood ADs and to promote their use when applicable, the Best Practices Subteam for DIA's Adaptive Design Scientific Working Group conducted an extensive nonsystematic search and reviewed trials from multiple sponsors who had employed these designs. Here, we review 10 specific case studies for which less well-understood ADs were employed and share feedback about their challenges and successes, as well as details about the regulatory interactions from these trials. We learned that these designs and associated statistical methodologies can make difficult research situations more amenable for study and, therefore, are needed in our toolbox. While they can be used to study many diseases, they are particularly valuable for rare diseases, small populations, studies involving terminal illnesses, and vaccine trials, in which it is important to find efficient ways to bring effective treatments to market more rapidly. It is imperative, however, that these methodologies be utilized appropriately, which requires careful planning and precise operational execution.

Keywords

adaptive trials, phase IIB/III trial designs, sample size re-estimation, data monitoring committee, 2010 FDA Adaptive Design guidance, adaptive dose selection

Introduction

Adaptive design (AD) clinical trials use accumulating subject data to modify the parameters of the design of an ongoing study, without compromising the validity and integrity of the study. They are generally used to move a compound through clinical development in an efficient manner. The 2010 Food and Drug Administration (FDA) Guidance for Industry: Adaptive Design Clinical Trials for Drugs and Biologics described a subset of 7 primary designs as "less well-understood" (ie, less experienced) due to relatively little experience in regulatory settings such that their performance properties were not fully understood at that time.[1,2] This notion of limited experience was reinforced in a 2010 report[3] in which a survey of 13 large- and medium-sized pharmaceutical companies identified only 59 AD clinical trials conducted between January 2003 and March 2008. However, since 2008, surveys conducted by the Center for Biologics Evaluation and Research (CBER), the Center for Devices and Radiological Health (CDRH), and the DIA's Adaptive Design Scientific Working Group (DIA ADSWG) all indicate increased implementation of ADs, including those in the "less experienced" category.[4,5] Barriers

[1] Independent biostatistical consultant, Levittown, PA, USA
[2] Statistical Methodology, Novartis Pharmaceuticals, East Hanover, NJ, USA
[3] Clinical Biostatistics, Merck & Co Inc, Rahway, NJ, USA
[4] Advanced Analytics Centre, AstraZeneca, Gaithersburg, MD, USA
[5] Center for Statistics in Drug Development, Quintiles Inc., Morrisville, NC, USA
[6] Amgen, Thousand Oaks, CA, USA
[7] Janssen Research and Development, Titusville, NJ, USA
[8] Pfizer Inc, Groton, CT, USA
[9] inVentiv Health Clinical, Ann Arbor, MI, USA
[10] Cytel, Cambridge, MA, USA

Submitted 13-May-2016; accepted 12-Jul-2016

Corresponding Author:
Eva Miller, PhD, 34 Fountain Road, Levittown, PA 19056, USA.
Email: evamiller425@msn.com

to adoption in pharmaceutical companies include concerns about regulatory acceptance; time involved in planning, communications, and teamwork; impact on randomization and drug supply management; extra considerations for independent data monitoring committees (DMCs); structuring and maintaining firewalls; and availability of appropriate software for planning, simulations, and implementation.[6] Despite the challenges, there remains a critical need for AD trials for efficient clinical development and minimizing patients' exposure to ineffective treatments. To better understand the qualities of these ''less experienced'' ADs and promote their use when applicable, the Best Practices Subteam for the ADSWG conducted a nonsystematic search; reviewed surveys conducted by CBER, CDRH, and Morgan; and spoke to their authors.[4,5,6] They reviewed trials from multiple sponsors who had employed these designs, assessed the challenges and successes of these ADs, and engaged in determining best practices for their future application.

The purpose of this manuscript is 2-fold: (1) to review specific studies in which ''less experienced'' ADs were used and to share details about regulatory interactions from these trials and (2) to acknowledge that, collectively, the scientific community has a more robust understanding of some designs within this subset, but for now, perhaps they will continue to be reviewed on a case-by-case basis.

Recent "Less Experienced" Adaptive Design Case Studies

For the purposes of this manuscript, the DIA ADSWG Best Practices Subteam selected 10 AD studies with sufficient detail available related to sponsor/agency interactions, challenges, and/or best practices developed as a result of the trial. In all cases, the studies have been classified as ''less experienced'' (Table 1) and have presented specific challenges related to their design. These challenges and the associated outcomes of the studies are summarized in Table 2. Here, we review these 10 trials.

ADVENT

ADVENT was a phase III, randomized double-blind, placebo-controlled, 2-stage adaptive dose-selection study evaluating the safety and efficacy of crofelemer to treat noninfectious diarrhea in human immunodeficiency virus (HIV) patients receiving antiretroviral therapy.[7] The primary efficacy endpoint was clinical response.

Two previous randomized, controlled trials showed the efficacy of crofelemer in other patient populations, using different dose levels and dose frequencies. Questions remained about the optimal treatment regimen for the treatment of diarrhea in HIV patients.

In stage 1 of the study, 200 patients were to be equally allocated to crofelemer oral doses of 125, 250, and 500 mg, twice daily, or to placebo. One of the active drug doses and placebo would be used in stage 2 with an additional 150 patients randomized, for a total study sample size of 350 patients. At the interim, the lowest dose that showed an improvement in clinical response of at least 2% was to be selected. A closed testing procedure and the inverse normal P value combination method assured strong control of type I error.[8] Extensive simulations evaluated type I error control, power considerations, and dose selection properties.

The study initiated in October 2007. In June 2009, enrollment was temporarily paused for 2 months to avoid enrolling patients to doses that would subsequently be discontinued while data for 194 evaluable subjects were collected, analyzed, and reviewed by the DMC. The 125-mg dose was selected for further testing. Its identity was restricted and was available on a need-to-know basis. Strict procedures prevented interim results from becoming public, thereby minimizing risk to trial integrity and validity. Enrollment resumed, and 180 patients were randomized by October 2010. Combining data from both study stages, the results for the primary efficacy endpoint were statistically significant (18% on crofelemer vs 8% on placebo; $P = .0096$).

Despite extensive planning, the difference in results across the stages was unanticipated and emphasizes the need for trial planners to consider temporal changes in the subject population and activated study sites. Additionally, this study illustrated the critical importance of analyzing interim data properly. An incorrect derivation of the primary endpoint resulted in the treatment effect being underestimated at the interim analysis, affecting the quality of the decision. Fortunately, the DMC was able to make the correct decision, leading to approval of an effective treatment.

VALOR

VALOR was an adaptive, 2-stage, phase III study evaluating safety and efficacy of vosaroxin as an add-on therapy to cytarabine in the treatment of patients with first relapsed or refractory acute myeloid leukemia.[9] Increased upfront planning was needed to fully understand study design operating characteristics. The sponsor engaged both FDA and European Medicines Agency (EMA) early on to secure agreement on the design. Strong documentation (simulation report, DMC charter, statistical analysis plan [SAP], and communication plan) was developed to ensure minimal operational bias. As the study results were borderline, EMA allowed the sponsor to continue with the marketing authorization, whereas FDA, reviewing the same information, did not, barring a second confirmatory study.

Initial power calculations required 375 events from 450 evaluable patients to achieve 90% power to detect a 40% survival improvement (5 vs 7 months median overall survival), corresponding to a 0.71 hazard ratio. Enrollment was planned for 24 months with 6 months' follow-up. The study design employed a Promising Zone approach[10] at the end of stage 1, after 50% of the events had been observed; at that time, a one-time 50% increase in the number of events and sample size could be implemented to repower the study for a weaker treatment effect. This adaptation depended on the interim

Table 1. Study Characteristics by Adaption Category.

Study Name[a]	Therapeutic Area	Indication	Name of Drug	Name of Comparator	Primary Endpoint	Rationale
Adaptation for dose selection						
ADVENT (NCT00547898)	Immunology	Symptomatic relief of diarrhea in HIV patients receiving antiretroviral therapy	Crofelemer (Fulyzaq)	Placebo	Clinical response	Dose selection in stage 1, to be studied in stage 2
MERCK-VACCINE (NCT00543543)	Oncology	Prevention of cervical and other cancers in the anogenital area caused by human papillomavirus types 6, 11, 16, 18, 31, 33, 45, 52, and 58	Human papillomavirus vaccine, nine-valent (GARDASIL 9)	Human papillomavirus vaccine, four-valent (GARDASIL)	Combined incidence of HPV 31/33/45/52/58-related high-grade cervical abnormalities (CIN 2/3), adenocarcinoma in situ, invasive cervical carcinoma, high-grade vulvar intraepithelial neoplasia, high-grade vaginal intraepithelial neoplasia, vulvar cancer, or vaginal cancer (starting after 7 months)	Dose selection in stage 1, to be studied in stage 2
INHANCE (NCT0046356)	Pulmonary/ respiratory	Treatment of chronic obstructive pulmonary disease for improvement of respiratory airflow	Indacaterol (Arcapta; Onbrez)	Tiotropium (Spiriva); formoterol (FORADIL AEROLIZER); placebo	Trough FEV1 at week 12	Dose selection in stage 1, to be studied in stage 2. The only clinical development question remaining was what dose to study in phase III. Used this adaptive dosing design to address this question.
STAMPEDE (NCT00268476)	Oncology	Treatment of high-risk localized or metastatic prostate cancer in men who are being treated for the first time with long-term androgen deprivation therapy (ADT) or androgen suppression	ADT + intravenous zoledronic acid (Zometa); ADT + zoledronic acid + docetaxel (Taxotere); ADT + zoledronic acid + celecoxib (Celebrex)	ADT as a standard; ADT + docetaxel (Taxotere); ADT + celecoxib (Celebrex)	Overall survival (final analysis); failure-free survival for interim	Continue treatments that show promise; discontinue treatments with insufficient evidence of activity. This is a platform study.
Adaptation of sample size based on interim-effect size estimates						
VALOR (NCT01191801)	Oncology	First relapsed or refractory acute myeloid leukemia	Vosaroxin (QINPREZO) + cytarabine (Cytosar-U)	Placebo + cytarabine (Cytosar-U)	Survival	uSSR

(continued)

Table 1. (continued)

Study Name[a]	Therapeutic Area	Indication	Name of Drug	Name of Comparator	Primary Endpoint	Rationale
A-HeFT (NCT00047775)	Cardiovascular	Treatment of moderate to severe heart failure in African American patients	Isosorbide dinitrate/ hydralazine hydrochloride (BiDil)	Placebo	Composite, including all-cause mortality as a component	uSSR (uncertainty about effect sizes and expected variation of primary endpoint)
Adaptation of patient population based on treatment-effect estimates						
CHAMPION (NCT01156571)	Cardiovascular	Platelet inhibition during percutaneous coronary intervention	Cangrelor (Kengreal)	Clopidogrel (Plavix)	Composite of death, myocardial infarction, or ischemia-driven revascularization within 48 hours	uSSR; futility analysis
Adaptation of multiple study design features in a single study						
GAZVYA (NCT01010061; NCT01998880; NCT02053610)	Oncology	In combination with chlorambucil chemotherapy for the treatment of previously untreated chronic lymphocytic leukemia	Obinutuzumab (Gazyva) + chlorambucil (Leukeran)	Chlorambucil (Leukeran) as a standard; rituximab (Rituxan) + chlorambucil as an alternative	Investigator-assessed progression-free survival	Innovative sequential stage-wise design was used to address multiple objectives.
RAPTOR (NCT01000961)	Nephrology	Management of nephropathic cystinosis in adults and children ages 6 years and older	Cysteamine bitartrate (Procysbi)-delayed release formulation	Cysteamine bitartrate (Cystagon)–immediate-release formulation	Peak white blood cell cystine levels	uSSR (product has orphan-drug designation and, therefore, the number of potential subjects is small); non-inferiority study
Infantile Proliferating Hemangioma Study (NCT01074437)	Hematology/ oncology	Infantile proliferating hemangioma requiring systemic therapy	Propranolol hydrochloride (Hemangeol)	Corticosteroids or placebo	Success/failure based on evolution of target infantile hemangioma from baseline to week 24 Success was defined as complete or nearly complete resolution of the target infantile hemangioma at week 24 vs baseline	Dose selection in stage 1, to be studied in stage 2; uSSR; futility analysis

uSSR, unblinded sample size re-estimation.
[a]Details of each study are searchable by National Clinical Trial (NCT) identification number on the Clinical Trials.gov website (http://clinicaltrials.gov).

probability of trial success (as measured by conditional power under the observed effect size) falling within the "promising zone." The study could also be stopped early for overwhelming evidence of efficacy or futility. A DMC was set up to review the unblinded interim data and make recommendations to the sponsor regarding study design adjustments.

In the summer of 2012, the DMC reviewed the interim data, and recommended the sample size be increased to 50% more events and 675 patients. Final study results were unblinded in late 2014. The prespecified primary analysis, a weighted log-rank statistic using the Cui-Hung-Wang method,[11] showed an estimated hazard ratio of 0.87 (6.1 vs 7.5 months median

Table 2. Summary of Challenges and Outcomes by Case Study.

Study Name	Challenges	Outcome
ADVENT	1. Error in interim analysis, which could have led to an incorrect decision on dose selection. This interim analysis was corrected when the study was completed and analyzed. 2. The interim and final analyses used different data sources, leading to different response rates. 3. Treatment difference statistically significant for stage 1 only.	Product was approved. The amount of time saved over a traditional development plan is unclear.
VALOR	1. Increased upfront planning was needed to fully understand study design operating characteristics. The sponsor engaged both FDA and EMA early on to secure agreement on the design. Strong documentation (simulation report, DMC charter, SAP, and communication plan) was developed to minimize operational bias.	Product was approved for market by EMA, but *P* value was .06, and a second study is planned in the US.
GAZYVA	1. Increased planning to address a large study population of previously untreated chronic lymphocytic leukemia patients with comorbidities.	Gazyva was the first medical product approved with FDA's Breakthrough Therapy designation.
RAPTOR	1. The initial uSSR indicated that 30 subjects were sufficient. A second uSSR was performed because of the discovery of a calculation error in the reporting of white blood cell cystine levels. The second uSSR indicated that a total sample size of 36 subjects was required to achieve 90% power for a test at the 1-sided $\alpha = 0.02104$.	Product was approved.
MERCK-VACCINE	1. Increased planning to maintain blinding of the study team. 2. Additional efforts to ensure the data from phase IIB and phase III could be combined. 3. Delays in phase III transition due to an unanticipated request by the FDA for an end-of-phase II meeting related to dose selection and phase III implementation.	Product was approved by the FDA. Operational efficiencies resulted in both time and resource savings with a phase II/III adaptive design.
INHANCE	1. Increased planning and resources compared to nonadaptive designs. 2. FDA did not prospectively agree to the dose selection plan because of concerns it might select inappropriately high doses.	The study successfully demonstrated efficacy of indacaterol. FDA review subsequently did not agree with the doses selected for definitive study. Additional studies were conducted, leading to approval of the 75-μg dose, which had not been evaluated in stage 2 of this study.
Infantile Proliferating Hemangioma Study	1. Numerous simulations required for operating characteristics of study. At the time of the interim analysis, 460 subjects had enrolled compared with 20-40 that had been anticipated. A decision was made to do the interim analysis instead of changing to a fixed design in order to maintain the uSSR option.	Study had a successful outcome, and drug development time was considerably shortened.
CHAMPION	1. Logistical challenges due to the large sample size and number of sites. 2. Extensive planning and simulations were needed to have the design approved by FDA and EMA.	The study stopped for futility after enrolling 8877 patients. Post hoc analyses suggested a separate study with an endpoint less dependent on biomarkers be used. The futility rule in this trial prevented the sponsor from committing further patients to the study, and resources were reassigned to a subsequent trial, leading to submission and eventually market approval.
A-HeFT	1. At each of the 2 planned interim analyses, the DMC observed trends favoring survival benefit. However, no stopping boundaries had been prespecified. Ultimately, the DMC recommended stopping the study. Recommend study protocols give some flexibility to DMCs, rather than rigidly adhering to uSSR rules.	Regulatory approval of the treatment in the tested population.
STAMPEDE	1. Simultaneous assessment of many treatments against a single control arm in a cost-efficient way. 2. Intent to incorporate new suitable treatments as new agents emerge after trial was initiated. 3. Statistical and significant operational preplanning to enable seamless add/discontinue of treatment arm during trial.	Study is ongoing to inform clinical practice, not to register new treatments. Two arms were discontinued due to lack of benefit, and 3 arms were added to incorporate new agents and change in populations.

DMC, data monitoring committee; EMA, European Medicines Agency; FDA, Food and Drug Administration; SAP, statistical analysis plan; uSSR, unblinded sample size re-estimation.

overall survival) and a *P* value of .06. A preplanned stratified analysis of the data, based on disease status, age, and geographic location, produced a hazard ratio of 0.83 and *P* value of .02.

The adaptive sample size increase was appropriately adjusted for an otherwise underpowered study. A larger increase in sample size could have provided clearer evidence of success, but a stratified or competing risk primary analysis to control for the transplantation postrandomization event would have been key to garnering clear statistical significance. This novel study design, in conjunction with an innovative funding mechanism, made an otherwise prohibitively expensive trial feasible. Investors agreed to fund the increase in sample size based on promising interim results for $25 million in return for an increased participation payment on future net sales of the drug product.

GAZYVA

The GAZYVA trial was a 2-stage study comparing obinutuzumab in combination with chlorambucil to chlorambucil alone and comparing rituximab in combination with chlorambucil to chlorambucil alone for the treatment of patients with previously untreated chronic lymphocytic leukemia.[12] This registration trial used an innovative stage-wise design to address multiple objectives. The goal of the trial was to establish a chemoimmunotherapy option for older patients with chronic lymphocytic leukemia and with comorbidities. The primary endpoint was progression-free survival. At stage 1, obinutuzumab-chlorambucil was compared to chlorambucil and rituximab-chlorambucil was compared to chlorambucil; at stage 2, obinutuzumab-chlorambucil was compared to rituximab-chlorambucil.

A stage-wise design has 2 preplanned interim analyses: the first interim analysis at the end of stage 1 with the objectives of testing for superiority of each combination to chlorambucil and testing for futility of obinutuzumab-chlorambucil versus rituximab-chlorambucil; the second interim analysis for early stopping for superiority of obinutuzumab-chlorambucil versus rituximab-chlorambucil; and the final analysis of superiority of obinutuzumab-chlorambucil versus rituximab-chlorambucil. To control for the type I error, the study first performed a global closed testing for any difference in the primary endpoint among the three study arms before proceeding to the pairwise comparisons in stage 1.[13,14]

The type I error was allocated to one interim and one final analysis in stage 2 according to an O'Brien-Fleming efficacy boundary with a Lan-DeMets alpha spending approach.[12] The global hypothesis of no treatment differences between the three treatment arms in progression-free survival was tested using data from stage 1. The pairwise comparisons in stage 1 were not to be performed unless the global hypothesis was rejected.

The sponsor proposed that analyses and results comparing obinutuzumab-chlorambucil to rituximab-chlorambucil at stage 1 would remain confidential by the DMC until the end of the trial, unless the stage 2 interim efficacy boundary was crossed. Stage 1 data comparing rituximab-chlorambucil with chlorambucil alone was blinded until all patients had been randomized and finished treatment. In the labeling update submission based on stage 2 results, the sponsor also provided *P* values from updated stage 1 progression-free and overall survival analyses for considerations. Since *P* values from stage 1 updated analyses did not have a multiplicity-adjusted significance level to compare for interpretation, they were not in the label. The sponsor reported discussions with FDA on study design to be helpful.[13] FDA viewed the evaluation as case specific.[15] Obinutuzumab was approved by the FDA in November 2013 in combination with chlorambucil and was the first medical product approved with FDA's Breakthrough Therapy designation.

RAPTOR

The RAPTOR trial compared two formulations of cysteamine bitartrate: a delayed-release formulation (Procysbi) to an immediate-release formulation (Cystagon) in patients with nephropathic cystinosis.[16] This disease often presents within the first year of life and, if left untreated, can destroy vital organs. The control medication (Cystagon) in this trial needs to be given every 6 hours, reliably, in order to be effective. Procysbi on the other hand, has a more liberal dosing schedule, making it easier to comply with the regimen. This case study provides an example of how unblinded sample size re-estimation (uSSR) was successfully employed and accepted in a single confirmatory study for an orphan indication.

The primary endpoint for RAPTOR was the peak white blood cell cystine levels after 12 hours for Procysbi and after 6 hours for Cystagon. Clinical efficacy and safety were primarily evaluated in one trial: a phase III, multicenter, open-label, randomized, active controlled, cross-over, non-inferiority study. A non-inferiority margin of 0.30 nmol 1/2 cystine/mg protein was chosen given that the goal of therapy was to maintain a cystine level depletion similar to that of Cystagon.[17] Due to uncertainties about the intrasubject variance estimate, a uSSR procedure with one IA was instituted. It was unblinded since the intra-subject variance was estimated for each of the treatments separately and then pooled assuming variance homogeneity. There was no intention to stop early.

Explicitly, the 2-stage uSSR procedure was as follows. Twenty subjects were enrolled, and the intrasubject variance was estimated. The new variance estimate was used to recompute the sample size, at the 1-sided $\alpha = 0.02104$. This level accounted for α inflation caused by uSSR to ensure strong α control at the 0.025 level.[18] If the resulting sample size was 30 or fewer, enrollment continued until 30 subjects were enrolled. Otherwise enrollment continued until a total of no more than 50 subjects in order to achieve 90% power.

This procedure indicated that a total sample size of 36 subjects was required. Forty-three subjects were randomized, because of the unexpected availability of patients, and treated. The mean peak white blood cell cystine level (least squares mean $\pm$ standard error of the mean) measured in the per-

protocol population (N = 38) of subjects treated with Cystagon was 0.54 ± 0.05, compared with 0.62 ± 0.05 for Procysbi. The mean difference was 0.08 with the upper limit of the 2-sided 95.8% confidence interval of 0.16, which was lower than the 0.3 non-inferiority margin, thus achieving the non-inferiority objective.

This was the only phase III trial conducted for this compound and was the primary basis for approval of Procysbi in the US and EU.

MERCK-VACCINE

The MERCK-VACCCINE trial compared the original Gardasil human papillomavirus (HPV) vaccine to the Gardasil 9 HPV vaccine in the prevention of cervical cancer.[19] This large adaptive, phase IIB/III outcome trial was successfully implemented, providing pivotal support for the registration of Gardasil 9.

The original quadrivalent HPV (4vHPV) vaccine is effective in preventing cervical and other cancers in the anogenital area caused by HPV types 6, 11, 16, and 18, which account for approximately 70% of cervical cancers. The goal of developing a 9-valent HPV (9vHPV) vaccine, by including 5 additional HPV types, was to cover approximately 90% of cervical cancers. Since earlier phase II studies showed antigen levels for some of the original HPV types had to be adjusted in the 9vHPV vaccine in order to maintain the same immune response, a phase IIB study was required to select a dose formulation for further investigation in phase III. A large outcome trial was necessary to demonstrate that the selected 9vHPV vaccine not only maintained high efficacy against the four original HPV types (based on noninferior antibody response), but also demonstrated superior efficacy to the 4vHVP vaccine in preventing high-grade cervical, vulvar, and vaginal disease associated with the additional 5 HPV types (the primary endpoint).

In order to shorten typical development time and make this potentially important vaccine available earlier, the feasibility of a phase IIB/III AD was explored. One potential drawback of a phase II/III design is that, in typical development programs, a pause between the end of phase II and initiation of phase III allows a fuller understanding and synthesis of the implications of the phase II findings, as well as important interactions with regulatory authorities. This pause helps ensure an adequate ''learning'' phase in the development program before initiating the ''confirmatory'' phase. For the 9vHPV vaccine, however, a long pause was not considered necessary because of its similarity to the 4vHPV vaccine, and the extensive knowledge and experience gained from the successful 4vHPV development program.

Another potential drawback of a phase II/III adaptive trial is that a statistical adjustment is typically required to ensure that the type I error rate for the final analysis is not inflated because dose selection in the phase II portion is based on observed data. Consequently, the statistical efficiency of the design depends on the trade-off between the required statistical penalty and the efficiency gained by combining data from both parts. In this case, however, selection of the 9vHPV vaccine dose level to carry forward to the phase III portion was based on immunogenicity response to the original 4 HPV types and safety data; no primary efficacy data were available at the time of the IA. This feature, along with the conservative nature of the exact test used for the primary efficacy outcome, suggested that continuing to follow subjects randomized (during the phase II portion) to the selected 9vHPV vaccine dose group and the 4pHPV vaccine control group throughout the phase III portion would increase statistical efficiency. It should be noted, however, that methodology is available that allows the dose selection to be based on the primary or a correlated endpoint and still control the type I error rate (see, eg, Wassmer[20]).

Operational efficiencies resulted in both time and resource savings with a phase II/III AD. The gap in time between the end of phase IIB and initiation of phase III was reduced, and both time and cost savings were realized on protocol development, review and approval by IRBs, and study site initiations. Fewer subjects and shorter follow-up times were required in the phase III portion, since phase IIB subjects assigned to the selected dose group and control group continued on study. Given these considerations and the special features of the 9vHPV vaccine development program, this approach was highly attractive. FDA and the sponsor engaged in dialogue early, and an end of phase II discussion ensured that FDA agreed with the dose selected for phase III.[21] This trial was a success and points to considering the appropriateness of these designs for vaccine trials, although FDA does conclude that decisions will be made on a case-by-case basis.[22] The study met all of its primary and secondary objectives, and Gardasil 9 was approved by FDA in December 2014.

INHANCE

The 2-stage phase III INHANCE study compared 4 doses of indacaterol (75, 150, 300, or 600 μg) to placebo, formoterol, or tiotropium in patients with chronic obstructive pulmonary disease.[23,24] An IA at the end of stage 1 led to selection of 2 indacaterol doses (150 and 300 μg) to continue into stage 2. Neither the FDA nor the sponsor saw the data at the time of the dose selection. The nature of the specific FDA concern is presented on page 17 (''by requiring observed advantage over the marketed active comparators, the plan might result in selecting inappropriately high doses''), and subsequently FDA viewed that this had occurred. As to why the sponsor proceeded with the algorithm in spite of the stated concerns, we wouldn't be in position to know, though we might speculate that their hypothetical scenario reviews led them to feel the algorithm would pick acceptable doses. Although the trial successfully demonstrated the safety and efficacy of indacaterol, FDA subsequently did not agree with the doses selected for stage 2, based on review of the shallow efficacy curves.[25] Additional studies were conducted, leading to approval by FDA of only the 75-μg dose, which had not been evaluated in stage 2.

Dose selection at the end of stage 1 was based on 2 endpoints, trough forced expiratory volume in the first second

(FEV1) on day 15 and the area under the curve of FEV1; FDA had advised the sponsor that they did not agree with the selection criteria.[25] The primary efficacy endpoint of the study at the end of stage 2 was trough FEV1 at week 12. The final analysis used Bonferroni division of alpha across the 4 initial dose groups (ie, 0.0125 two-sided tests). Accounting for dropped doses in the testing procedure provided little improvement in power, while creating testing strategy difficulties for the multiple secondary endpoints. The sample size of 400 patients per group provided nearly 100% power for the primary endpoint of comparison to placebo and 85% power for the secondary objective of comparing indacaterol to tiotropium.

Prior clinical studies had determined appropriate choices of dosing schedule, inhaler device, patient population, and endpoints. The single critical design uncertainty was the precise dose level to study in phase III, promoting selection of an AD for the study because it was expected to reduce the time to complete the study and require fewer patients compared to a dose-finding phase II and a fixed-design phase III study, in spite of the increased planning and resources that were required. Trial simulations evaluated stage 1 sample size effects on the power of the complete study, optimal dose selection probabilities, and the overall sample size.

Public FDA documents indicated that FDA did not prospectively agree to the dose selection plan based on concern that, by requiring observed advantage over the marketed active comparators, the plan might result in selecting inappropriately high doses.[25] This was, in fact, FDA's judgment of the INHANCE results, requiring further investigation of the dose response via additional studies. Although the INHANCE study demonstrated the efficacy of indacaterol, and EMA did approve the doses selected in INHANCE, the FDA decision ultimately was to approve only a dose level lower than those brought forward to the stage 2 portion of the trial. Thus, the use of the adaptive design did not facilitate FDA approval of the investigational treatment in this program. Approval of a dose not promoted in the adaptive trial was only granted after more time, resources, and costs (ie, the additional trials) had been expended.

This case illustrates the importance when considering such designs of having high confidence by all parties involved regarding decision rules for IAs. As there will not be time for thoughtful consideration of the data, as might take place in the ''white space'' between separate phase II and phase III trials in a conventional program, these decision rules will enable the completed study to provide the information necessary for most effectively advancing the overall development program.

Infantile Proliferating Hemangioma

This study was a 2-stage phase II/III design comparing 3 active doses of propranolol (1, 2, and 4 mg/kg/d) to an active comparator of corticosteroids or placebo in patients with infantile proliferating hemangioma (IH) requiring systemic therapy. The primary endpoint was success or failure based on change of the target IH from baseline to week 24. This IH study is one of the first examples of a successful confirmatory AD, bringing propranolol to market for this orphan indication.

The trial objectives were 2-fold: (1) to demonstrate superiority of 1 or 2 propranolol regimens over placebo at the end of stage 2 and (2) to show that the selected regimen is safe in IH infants.[26] An IA was planned at the end of stage 1 (phase II), allowing for 3 possible adaptations: (1) selection of 1 or 2 active treatment regimens for further study in stage 2 (phase III), (2) uSSR, and (3) early stopping for futility, but not efficacy. The IA was conducted by an independent DMC.

One of the contributions that this study makes to our knowledge base is the sponsor's and authors' willingness to share details of what they learned in the regulatory interactions. Both EMA and FDA provided feedback on the study design. Not only did EMA request modification of the primary endpoint, but they suggested that the design would be improved by studying 2 dose/duration combinations instead of a single duration for three doses. They also required the IA to be based on the same time point as the primary analysis and more patients to be enrolled prior to the IA to be able to better study the regimens. Furthermore, they rejected the active comparator arm because corticosteroid was not an approved IH treatment in all EU countries. Consistent with EMA, FDA requested a modification of the primary endpoint. They also required a clearly delineated and detailed DMC charter and finalized SAP prior to study start. Although modifications to the DMC charter and the SAP could be made until treatment unblinding for the IA, FDA wanted to review, prior to study initiation, quasi-final versions of the charter and the interim and primary analysis sections of the SAP (including the results of an extensive simulation study), to ensure the validity and integrity of the trial. In addition, FDA recommended less aggressive futility stopping rules, prespecified guidelines on dose selection, and gatekeeping procedures to adjust testing of key secondary endpoints for multiplicity.[27]

In general, this study reinforces the importance of simulations and planning in order to carefully time the IA. One benefit of this planned AD was that it had the ability to discontinue poor-performing treatment arms at the IA. Its primary advantage, however, was that it reduced the overall development time by combining phases II and III into a single trial, given the unmet medical need for first-line treatment for proliferating IH with a good benefit-risk profile.

CHAMPION

The CHAMPION program included 3 phase III trials, of which the second one, PLATFORM, was a randomized double-blind, multicenter adaptive enrichment trial to study cangrelor in patients undergoing percutaneous coronary intervention.[28] The primary endpoint was a composite of death, myocardial infarction, or ischemia-driven revascularization within 48 hours. The study design was very sensitive to assumptions about the low event rates and the small treatment effect. Committing a large number of patients to a study with so much uncertainty was

risky. An AD built in safeguards in case observed data showed that planning assumptions were incorrect.

The design called for 8750 patients to be enrolled. Two IAs were planned allowing early trial termination due to efficacy or futility. The second IA, conducted after 70% of the subjects had observed data, provided an opportunity to increase the sample size (up to 16,000 patients) based on conditional power considerations and to enrich the study population. Enrichment was based on prespecified nested subpopulations. Strong control of type I error was ensured by applying a closed testing procedure and the inverse normal P value combination method.[29] Power considerations and enrichment properties were studied via simulation.

Extensive planning and simulations were needed to have the design approved by FDA and EMA. Early interactions with regulatory authorities were key, and logistical challenges were important because of the large sample size and number of sites. Ultimately, the study stopped for futility after having enrolled 8877 patients. However, post hoc analyses suggested a separate study with an endpoint less dependent on biomarker-defined myocardial infarction be used.

A follow-up study, PHOENIX, was performed with a composite endpoint that also included stent thrombosis and a different definition of myocardial infarction.[30] PHOENIX included an option to re-estimate the sample size, although that did not take place. Cangrelor was found to be superior to clopidogrel in this new trial, leading to submission and eventually market approval. Ultimately, the most useful adaptation was the futility rule in PLATFORM that saved the sponsor from committing further patients to a trial studying the wrong endpoint. Resources were reassigned to the PHOENIX trial, leading to a successful outcome.

A-HeFT

The A-HeFT trial compared isosorbide dinitrate/hydralazine hydrochloride (BiDil) to placebo in African American heart failure patients.[31] The primary endpoint involved a novel composite score algorithm, including all-cause mortality as a component. The primary endpoint and its mortality component reached statistical significance, resulting in regulatory approval for the treatment.

The initial sample size was 800 patients and two IAs were planned. Because of the novelty of the endpoint, there was uncertainty about the effect size and the variability. Thus it was planned to perform a uSSR at the second IA using the Cui-Hung-Wang method.[11] The DMC agreed to implement this and informed the Steering Committee about a revised sample size.

At the first IA, the DMC noticed a mortality trend favoring BiDil. No mortality stopping boundaries had been prespecified, but on ethical grounds they implemented a spending function scheme to guide their continuing mortality review. By the second IA, 528 patients were enrolled. The SSR algorithm yielded a revised target of 1100 patients, 300 more than the original design. The mortality advantage had increased, and in fact fell just below the boundary that had been defined. The DMC desired an additional review and felt there was a good chance to later terminate the study for overwhelming mortality benefit. This induced a conflict, as they had agreed to implement the SSR that would lead to a substantial sample size increase, yet the study might be trending toward termination for mortality. They ultimately recommended the sample size increase as per the agreed-upon algorithm, but requested an additional review several months later. At the time of the review, enrollment had reached 1050. The mortality hazard ratio of 0.57 favoring BiDil reached the boundary. The DMC recommended termination, and the study was stopped.

Sample size re-estimation schemes are commonly defined based on the primary endpoint. Data Monitoring Committees review the totality of information available, including other important efficacy and safety outcomes, so a decision may be more complex than simply focusing on the main endpoint. Study planning might have acknowledged that unexpected circumstances could lead the DMC to deviate from the SSR scheme, and the plan could have been stated in a manner that would enable discretion to be used if needed. Sample size re-estimation methods generally do not require rigid algorithmic change; they simply enable the possibility, allowing valid final analyses even if sample size remains unchanged. A slight protocol wording change might have allowed the DMC some flexibility to utilize their expertise in deciding how to react to unusual circumstances.

A-HeFT is an excellent example of a subgroup study and of use of a novel primary endpoint. The trial results quickly reflected the improved mortality rates in the self-identified African American population. It may have been very helpful if the DMC charter had explicitly allowed the committee members more discretion in the details of the implementation of the uSSR algorithm.

STAMPEDE

STAMPEDE is an ongoing open-label, randomized, controlled trial in prostate cancer using multiarm, multistage (MAMS) design, with the aim to inform clinical practice by combining standard androgen deprivation therapy (ADT) with one or more approved agents. This early use of MAMS addressed many practical issues of methodology.[32] The study was initiated in 2005 with 6 arms: ADT alone as standard of care and in combination with zoledronic acid, docetaxel, celecoxib, zoledronic acid with docetaxel, or zoledronic acid with celecoxib. Twice as many patients were allocated to ADT alone as any other arm to improve the estimate of the control rate used for comparison to each other arm.

Failure-free survival was used as an intermediate outcome in the three IAs to determine whether an arm should be terminated. The IAs were prospectively planned to use a hazard ratio criterion for treatment retention that progressed from liberal to more stringent to avoid discontinuing a treatment too early. The study had a planned specified end with use of overall survival as the final endpoint, and each remaining combination

treatment was compared to the ADT-only group with alpha of 0.05.[32] Strong control of the type I error rate was not implemented as the study was planned to inform clinical practice and not for regulatory approval purposes. For regulatory approval, the MAMS set-up can be used with appropriate adjustment for multiplicity (see, eg, Wassmer[20]).

The MAMS design allows for the simultaneous assessment of several treatments against a single control arm, a more efficient approach than individual trials. The MAMS design of comparing each investigational group to only the common control also avoided the potential for interactions between drugs (either favorable or unfavorable) to obscure benefits, which might occur in standard factorial design analyses. Additional efficiency was gained as, unlike in a factorial design, unpromising treatment groups could be terminated at an IA.

The MAMS design, an example of a platform trial, also permits flexibility to incorporate new treatment arms into the ongoing study when external information indicates their potential. The flexibility of study revision coupled with careful operational planning was applied multiple times[33–35]: 2 arms were dropped, 3 arms completed recruitment which led to a change in standard of care, and 3 new arms were sequentially added.

Although not a regulatory-rigorous study, STAMPEDE illustrates the value of adapting the design of an ongoing study based on either internal or external information to achieve efficiency, and that the study objectives will influence the balance of rigor with design flexibility.

Discussion

In this paper the DIA ADSWG Best Practices Subteam reviewed a number of ADs with features classified as ''less-well understood'' according to the 2010 FDA Draft Guidance.[1] Table 1 summarizes the study characteristics and adaptation categories for each case study.

By examining these case studies, we hoped to learn more about the challenges and successes of these trials. This raises the question, when is an AD actually considered successful? The definition of ''success'' needs to accommodate that ADs might be used in phase II development to efficiently answer questions critical to the design of subsequent phase III studies, or in phase III studies to answer a limited amount of uncertainty that impedes efficient use of a fixed design to provide the strength of evidence to support the intended regulatory objectives. An AD is still valuable when it leads to an efficient study that directly or indirectly supports marketing approval, or when it demonstrates that a therapy is not viable. Perhaps a better defining question is, Did the study accomplish its stated objectives: both the immediate quantitative data design question and the objective of advancing the development program to the intended next stage? It is a more complex question, but it recognizes that the best use of AD studies will be an important part of a good development program. How large a part, and what part, will vary from drug program to drug program.

We agree that a ''successful'' AD trial requires an assessment of whether the AD study was designed and implemented in an intelligent way. If the study provides answers to the wrong questions for the development program, then that study is not a success by this definition.

Compared to traditional trials, more attention is required in the planning stage. As we have seen with, for example, A-HeFT, we must ''expect the unexpected.'' Communications with the regulatory agency prior to formal submission could be an efficient approach to assess the appropriateness of conducting an AD study. Detailed SAPs for the complete study and for each IA may be needed before the study starts. Extensive simulation studies are greatly helpful in demonstrating the control of type I error rate for complex designs.[4] Stopping rules and success criteria, including specific IA decision rules, should be prespecified and documented. Considerations for preserving study integrity, including complete documentation of firewalls to minimize operational bias, are essential. The role of the DMC for an AD requires more thought and clear documentation in a DMC Charter than for a traditional trial. As we learned from a number of these trials, (1) validated data combined with accurate analyses are required for IAs; (2) the appropriate timing of each IA, related to enrollment rates, number of events, and variability, is crucial to the success of the trial and the accuracy and efficiencies to be achieved through utilization of ADs.

It is worth noting that, even with careful planning, AD trials may fall beyond the scope of anticipated ''unexpected'' situations/scenarios, and the interpretability of these study results will be in question, especially when the studies involve multiple adaptation features. Nonetheless, it may be worth the effort to implement ADs if one believes the advantages/benefits outweigh the challenges.

Conclusions

We learned that these ''less experienced'' ADs and associated statistical methodologies can make difficult research situations more amenable for study and, therefore, are needed in our toolbox. With the growing experience from the case studies and other AD trials in recent years, we feel that the scientific community has a more robust understanding of complex ADs and some ''less well-understood'' designs are becoming ''more understood.'' It is imperative, however, that AD methodologies be utilized appropriately, which requires careful planning and precise operational execution. The DIA ADSWG continues to work with sponsors and regulatory authorities to gain greater experience and acceptance of these methods, so that they will be used appropriately among trial designers, trial implementers, and regulatory authorities.

Author Note

The views and opinions expressed in this article are those of the individual authors and should not be attributed to any agency or organizations with which the authors are employed or affiliated.

Acknowledgments

We acknowledge the DIA ADSWG Best Practices Subteam members for their contributions to data collection, review, and discussion of trials utilizing less-well understood adaptive designs. In addition, we acknowledge Noelle Gasco (inVentiv Health Clinical) for her contributions to the editorial review of this document.

Declaration of Conflicting Interests

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: Eva Miller owns shares of Johnson & Johnson. Weili He and Kenneth Koury own shares of Merck & Co. Qi Jiang is an employee and shareholder of Amgen. Marc Walton is an employee of Janssen. Cunshan Wang is an employee and shareholder of Pfizer. Katherine Woo is an employee and shareholder of Johnson & Johnson. Paul Gallo is an employee of Novartis. Lisa Kammerman is an employee of AstraZeneca.

Funding

No financial support of the research, authorship, and/or publication of this article was declared.

References

1. US Food and Drug Administration. *Guidance for Industry: Adaptive Design Clinical Trials for Drugs and Biologics*. Washington, DC: US Department of Health and Human Services; 2010.
2. He W, Gallo P, Miller E, et al. Addressing challenges and opportunities of "less well-understood" adaptive designs. *Therapeutic Innovation & Regulatory Science*. 2016.
3. Quinlan J, Gaydos B, Maca J, Krams M. Barriers and opportunities for implementation of adaptive designs in pharmaceutical product development. *Clin Trials*. 2010;7:167-173.
4. Lin M, Lee S, Zhen B, et al. CBER's experience with adaptive design clinical trials. *Therapeutic Innovation & Regulatory Science*. 2016;50(2):195-203.
5. Yang X, Thompson L. Adaptive design practice at CDRH, January 2007–May 2013. Paper presented at JSM 2015.
6. Morgan C, Huyck S, Jenkins M, et al. Adaptive design: Results of 2012 survey on perception and use. *Therapeutic Innovation & Regulatory Science*. 2014;48:473-481.
7. Chaturvedi PR, Antonijevic Z, Mehta C. Practical considerations of a two-stage confirmatory adaptive clinical trial design and its implementation: ADVENT trial. In: He W, Pinheiro J, Kuznetsova OM, eds. *Practical Considerations for Adaptive Trial Design and Implementation*. New York, NY: Springer; 2014: 383-411.
8. Posch M, Koenig F, Branson M, Brannath W, Dunger-Baldauf C, Bauer P. Testing and estimation in flexible group sequential designs with adaptive treatment selection. *Stat Med*. 2005;24: 3697-3714.
9. Ravandi F, Ritch E, Sayar H, et al. VALOR, an adaptive design, pivotal phase 3 trial of vosaroxin or placebo in combination with cytarabine in first relapsed or refractory acute myeloid leukemia. http://www.sunesis.com/data-pdf/595/sunesis-valor-vosaroxin-201206-ASCO.pdf. Accessed May 27, 2015.
10. Mehta CR, Pocock SJ. Adaptive increase in sample size when interim results are promising: A practical guide with examples. *Stat Med*. 2011;30:3267-3284.
11. Cui L, Hung HM, Wang SJ. Modification of sample size in group sequential clinical trials. *Biometrics*. 1999;55:853-857.
12. Goede V, Fischer K, Busch R, et al. Obinutuzumab plus chlorambucil in patients with CLL and coexisting conditions. *N Engl J Med*. 2014;370:1101-1110.
13. US Food and Drug Administration, Center for Drug Evaluation and Research. Application number: 125486Orig1s000, Statistical Reviews. BLA/serial number: BLA 125486/00. http://www.accessdata.fda.gov/drugsatfda_docs/nda/2013/125486Orig1s000StatR.pdf. Completed August 29, 2013. Accessed October 15, 2015.
14. Ko C-W, Nie L, Sridhara R. Sequential stage design in the GAZYVA Registration Trial. Presented at the FDA Center for Drug Evaluation and Research Office of Biostatistics/Division of Biometrics 5, Joint Adaptive Design and Bayesian Statistics Conference; February 11-12, 2015; Washington DC.
15. US Food and Drug Administration, Center for Drug Evaluation and Research. Application number: 125486Orig1s000, Medical Reviews. http://www.accessdata.fda.gov/drugsatfda_docs/nda/2013/125486Orig1s000MedR.pdf. Completed October 1, 2013. Accessed October 15, 2015.
16. Langman CB, Greenbaum LA, Sarwal M, et al. A randomized controlled crossover trial with delayed-release cysteamine bitartrate in nephropathic cystinosis: effectiveness on white blood cell cystine levels and comparison of safety. *Clin J Am Soc Nephrol*. 2012;7:1112-1120.
17. Dohil R, Fidler M, Gongoiti JA, Kaskel F, Schneider JA, Barshop BA. Twice-daily cysteamine bitartrate therapy for children with cystinosis. *J Pediatr*. 2010;156:71-75.
18. Friede T, Kieser M. Sample size recalculation in internal pilot study designs: a review. *Biom J*. 2006;48:537-555.
19. Joura E, Giuliano A, Iversen O, et al. A 9-valent HPV vaccine against infection and intraepithelial neoplasia in women. *N Engl J Med*. 2015; 372:711-723.
20. Wassmer G. On sample size determination in multi-armed confirmatory adaptive designs. *J Biopharm Stat*. 2011;21:802-817.
21. Chen Y, Gesser R, Luxembourg A. A seamless Phase IIB/III adaptive outcome trial: design rationale and implementation challenges. *Clin Trials*. 2015; 12:84-90.
22. Lin M, Lee S, Zhen B, et al. CBER's experience with adaptive design clinical trials. *Therapeutic Innovation & Regulatory Science*. 2016;50(2):195-203.
23. Barnes PJ, Pocock SJ, Magnussen H, et al. Integrating indacaterol dose selection in a clinical study in COPD using an adaptive seamless design. *Pulmon Pharmacol Ther*. 2010;23: 165-171.
24. Donohue JF, Fogarty C, Lötvall J, et al. Once-daily bronchodilators for chronic obstructive pulmonary disease: indacaterol versus tiotropium. *Am J Respir Crit Care Med*. 2010;182:155-162.
25. Durmowicz AG. September 29, 2009 cross-discipline team leader review. In: Arcapta Neohaler FDA medical review document. Food and Drug Administration website. http://www.accessdata.

fda.gov/drugsatfda_docs/nda/2011/022383Orig1s000MedR.pdf. Accessed April 19, 2016.

26. Léauté-Labrèze C, Hoeger P, Mazereeuw-Hautier J, et al. A randomized, controlled trial of oral propranolol in infantile hemangioma. *N Engl J Med*. 2015;372:735-746.
27. Heritier S, Morgan-Bouniol CC, Lo S, Gautier S, Voisard JJ. A single pivotal adaptive trial in infants with proliferating hemangioma: rationale, statistical challenges, experience and recommendations. In: Sverdlov O, ed. *Modern Adaptive Randomized Clinical Trials: Statistical, Operational, and Regulatory Aspects*. New York, NY: Chapman & Hall/CRC Press; 2015:453.
28. Harrington RA, Stone GW, McNulty S, et al. Platelet inhibition with cangrelor in patients undergoing PCI. *N Engl J Med*. 2009; 361:2318-2329.
29. Mehta CR, Gao P. Population enrichment designs: case study of a large multinational trial. *J Biopharm Stat*. 2011;21:831-845.
30. Bhatt DL, Stone GW, Mahaffey KW, et al. Effect of platelet inhibition with cangrelor during PCI on ischemic events. *N Engl J Med*. 2013;368:1303-1313.
31. Taylor AL, Ziesche S, Yancy C, et al. Combination of isosorbide dinitrate and hydralazine in blacks with heart failure. *N Engl J Med*. 2004;351:2049-2057.
32. Sydes MR, Parmar MK, James ND, et al. Issues in applying multi-arm multi-stage methodology to a clinical trial in prostate cancer: the MRC STAMPEDE trial. *Trials*. 2009;10:39.
33. Sydes MR, Parmar MK, Mason MD, et al. Flexible trial design in practice—stopping arms for lack-of-benefit and adding research arms mid-trial in STAMPEDE: a multi-arm multi-stage randomized controlled trial. *Trials*. 2012;13:168.
34. Parker CC, Sydes MR, Mason MD, et al. Prostate radiotherapy for men with metastatic disease: a new comparison in the Systemic Therapy in Advancing or Metastatic Prostate Cancer: Evaluation of Drug Efficacy (STAMPEDE) trial. *BJU Int*. 2013;111: 697-699.
35. Attard G, Sydes MR, Mason MD, et al. Combining enzalutamide with abiraterone, prednisone, and androgen deprivation therapy in the STAMPEDE trial. *Eur Urol*. 2014;66: 799-802.