

# Challenges and solutions in the design and execution of the PROSPECT Phase II/III neoadjuvant rectal cancer trial (NCCTG N1048/Alliance)

Deborah Schrag<sup>1</sup>, Martin Weiser<sup>2</sup>, Leonard Saltz<sup>3</sup>, Harvey Mamon<sup>4</sup>, Marc Gollub<sup>5</sup>, Ethan Basch<sup>6</sup>, Alan Venook<sup>7</sup> and Qian Shi<sup>8</sup>

## Abstract

**Background:** More than half of the 40,000 incident rectal cancer patients in the United States each year are diagnosed at clinical stage II and III (locally advanced stage). For this group, high rates of cure can be achieved with the combination of pelvic radiation and sensitizing 5-fluorouracil (chemoradiation), surgery and chemotherapy, but treatment is long, arduous and toxicities are substantial. The PROSPECT trial (N1048, NCT01515787) was designed to determine whether neoadjuvant chemotherapy with 5-fluorouracil and oxaliplatin (FOLFOX) could be used as an alternative to neoadjuvant chemoradiation without compromising treatment outcomes and to spare these patients excess toxicity. The statistical design balanced the twin co-primary goals of achieving low local and distant recurrence rates. Study design features contended with the need for stringent safeguards given limited phase II data, the need for straightforward criteria to facilitate both accrual and protocol fidelity and the importance of patients' perspectives on symptom burden and treatment toxicity.

**Methods:** PROSPECT is an ongoing multi-site two-group seamless phase II/III randomized trial comparing standard neoadjuvant chemoradiation versus neoadjuvant chemotherapy with selective use of chemoradiation for patients with locally advanced rectal cancer. Challenges addressed in the design and conduct of PROSPECT have included the following: (1) setting safety thresholds given limited single-center phase II data, (2) establishing workable eligibility criteria, (3) balancing competing time to local and distant recurrence as co-primary endpoints and (4) obtaining reliable and complete data for patients' symptom burden. The design and implementation challenges, choices, modifications and their implications for the design of future national cooperative group clinical trials are presented.

**Results:** PROSPECT incorporated stringent thresholds for both complete surgical resection (R0) and the time to local recurrence as early stopping rules. When predetermined stopping criteria were not met after evaluation of the first 366 participants in the randomized phase II, the study transitioned seamlessly to phase III with cumulative accrual of over 1000 participants. Eligibility criteria stipulating rectal tumor location based on distance from the anal verge were unworkable, and the protocol was amended to a more pragmatic approach that assigned surgeons with primary responsibility for determining eligibility. Central radiology review was feasible and in some cases prompted discontinuation of protocol treatment. Participation in toxicity reporting using the National Cancer Institute's Patient-Reported Outcomes Version

<sup>1</sup>Division of Population Sciences, Dana-Farber Cancer Institute and Harvard Medical School, Boston, MA, USA

<sup>2</sup>Department of Surgery, Memorial Sloan Kettering Cancer Center and Weill Medical College of Cornell University, New York, NY, USA

<sup>3</sup>Department of Medical Oncology, Memorial Sloan Kettering Cancer Center and Weill Medical College of Cornell University, New York, NY, USA

<sup>4</sup>Department of Radiation Oncology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA

<sup>5</sup>Department of Radiology, Memorial Sloan Kettering Cancer Center and Weill Medical College of Cornell University, New York, NY, USA

<sup>6</sup>Division of Medical Oncology, UNC Lineberger Comprehensive Cancer Center and University of North Carolina School of Medicine, Chapel Hill, NC, USA

<sup>7</sup>Division of Medical Oncology, University of California San Francisco Cancer Center and University of California San Francisco School of Medicine, San Francisco, CA, USA

<sup>8</sup>Alliance Statistics and Data Center, Mayo Clinic Cancer Center, Rochester, MN, USA

## Corresponding author:

Deborah Schrag, Division of Population Sciences, Dana-Farber Cancer Institute and Harvard Medical School, 450 Brookline Avenue, D-1114 Boston, MA 02215, USA.

Email: deb\_schrag@dfci.harvard.edu

of the Common Terminology Criteria for Adverse Events was uniformly high and was well accepted by participants from over 200 sites in the United States, Canada and Switzerland.

**Conclusion:** The strategies used to overcome these obstacles may inform the design of other studies that involve multi-modality treatment interventions, particularly trials where implementation of consistent criteria for eligibility and outcomes across hundreds of practice settings is necessary.

### Keywords

Locally advanced rectal cancer, neoadjuvant FOLFOX, total mesorectal excision, combined modality treatment, chemoradiation, patient-reported outcomes and symptom burden using Patient-Reported Outcomes Version of the Common Terminology Criteria for Adverse Events, central radiology review of pelvic magnetic resonance imaging, phase II/III hybrid design, co-primary endpoints, pragmatic design elements in cooperative group cancer clinical trials

## Introduction, background and rationale

A 5.5-week course of pelvic radiation with sensitizing 5-fluorouracil (5FU) (chemoradiation) has been a cornerstone of treatment for locally advanced (stage II and III) rectal cancer for 30 years.<sup>1,2</sup> In 2004, a landmark German study established that preoperative administration of chemoradiation achieved a lower rate of local recurrence and superior quality of life but no survival advantage compared to postoperative chemoradiation.<sup>3</sup> Subsequently, preoperative chemoradiation became the standard approach for management of stage II (T3/T4N0) and stage III (TanyN1/N2) rectal cancers worldwide.

Key aspects of rectal cancer therapeutics have evolved since those paradigm-changing events. Advances in imaging, surgical technique and chemotherapy called into question whether trimodality approaches are essential.<sup>4–8</sup> The current standard rectal cancer treatment timeline is 5.5 weeks of chemoradiation, a 4- to 6-week recovery, surgical resection, another 4- to 6-week recovery and then adjuvant therapy. Because no systemic therapy is delivered for over 3 months, there is a window for early dissemination—another rationale for moving chemotherapy earlier in the treatment timeline.

For these reasons, investigators at Memorial Sloan Kettering Cancer Center conducted a single-institution pilot single-group phase II trial of 32 patients that suggested it was safe to selectively omit chemoradiation for rectal cancer patients who had evidence of clinical response to neoadjuvant FOLFOX chemotherapy.<sup>9</sup> The favorable outcomes including a low rate of local failure and high pathologic complete response rate in this pilot trial led investigators in the North Central Cancer Treatment Group to design PROSPECT (N1048), a multicenter phase II/III NCI cooperative group trial that has recruited its target of 1194 participants. The North Central Cancer Treatment Group is now part of the Alliance for Clinical Trials in Oncology.

PROSPECT investigators confronted at least four key challenges: (1) setting safety thresholds given

limited single-center phase II data, (2) establishing workable eligibility criteria and safeguards, (3) balancing competing time to local and distant recurrence as co-primary endpoints and (4) eliciting patients' direct reports of treatment adverse effects. This report describes key design challenges and strategies employed to overcome them.

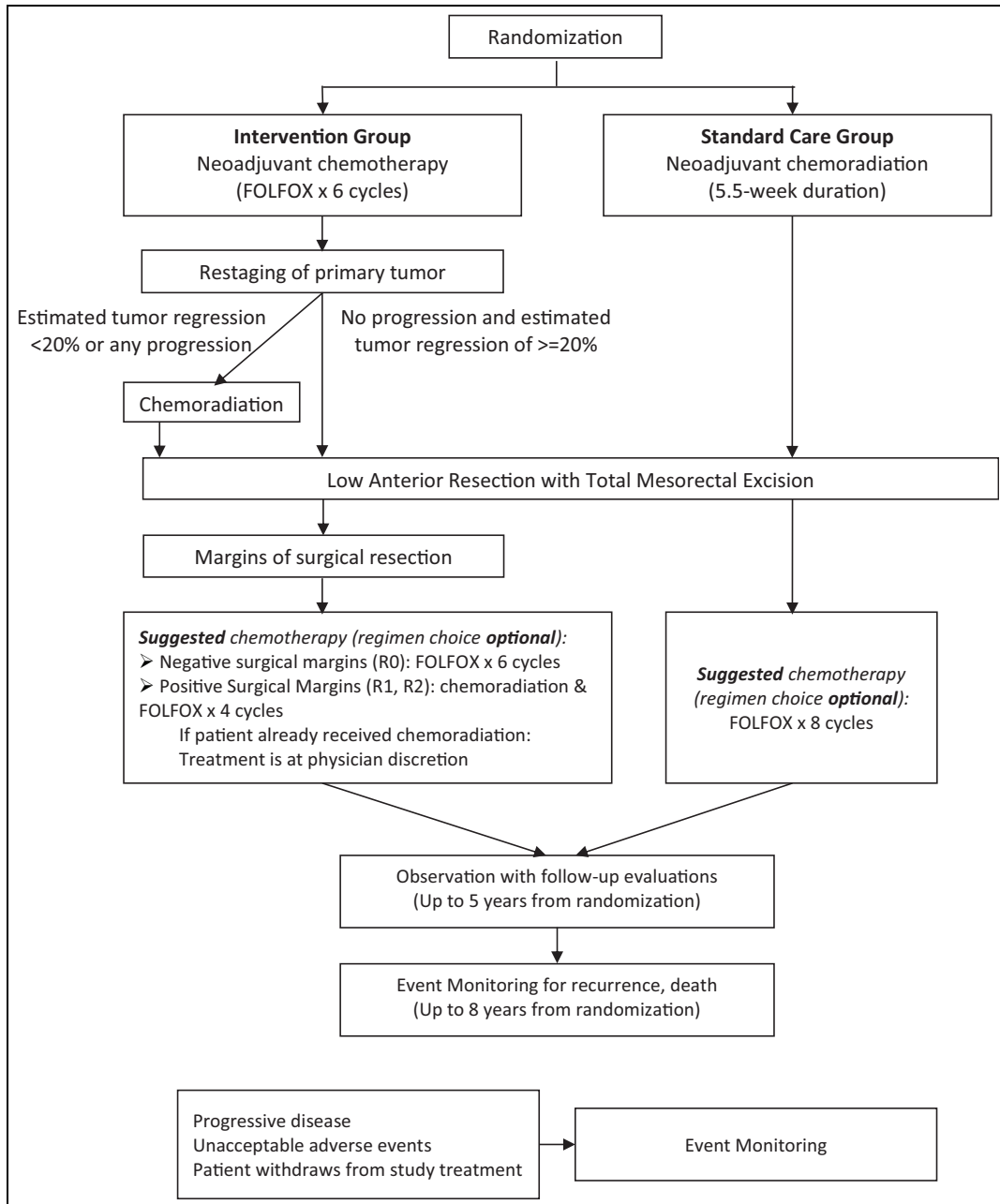
## Methods

PROSPECT is a multicenter phase II/III randomized controlled trial to determine whether six cycles of neoadjuvant FOLFOX followed by comprehensive restaging with selective use of pelvic chemoradiation achieves favorable outcomes for patients with locally advanced (clinical T3N0, T3N1, T2N1) rectal cancer in comparison with the standard use of preoperative pelvic chemoradiation prior to a total mesorectal excision (TME). Figure 1 shows the study schema.

Eligible study subjects include adults diagnosed with rectal adenocarcinoma of clinical stage T2N1, T3N0 or T3N1 who are candidates for curative intent sphincter-sparing surgery and lack high-risk features such as tumor encroaching upon the mesorectal fascia or distal tumors (Table 1). Patients from 276 participating sites are randomized 1:1, stratified by ECOG (Eastern Cooperative Oncology Group) performance status—(0 and 1) versus 2.

In the intervention group, participants receive six cycles of neoadjuvant FOLFOX followed by repeat tumor imaging and proctoscopy. Clinical estimation is made as to whether there has been at least a 20% decrease in the tumor in response to neoadjuvant FOLFOX. Participants with  $\geq 20\%$  response proceed to surgery with TME. Participants with  $< 20\%$  response receive neoadjuvant chemoradiation and then proceed to surgery with a TME. In the standard group, participants undergo TME surgery after chemoradiation without interim restaging.

Each participant signs an institutional review board (IRB)-approved, protocol-specific informed consent



**Figure 1.** Protocol schema.

document in accordance with federal and institutional guidelines. Data collection and statistical analyses are conducted by the Alliance Statistics and Data Center. Data quality is ensured by review of data by the Alliance Statistics and Data Center and by the study chairperson following Alliance policies.

### **Challenge #1: limited phase II preliminary data**

When PROSPECT was proposed, evidence supporting safety and favorable long-term outcomes when omitting chemoradiation was limited. Clinical trialists had trepidation about the safety of the investigational approach

across hundreds of centers with different care models and where care is not typically provided by a team of colorectal subspecialists.

### **Challenge #2: establishing eligibility and safety criteria that could be consistently applied across sites**

Investigators sought to incorporate stringent criteria to ensure safety of participants randomized to the intervention group in the multicenter context, specifically, surgeon credentialing and central radiology review. First, the initial protocol stipulated that participating

**Table 1.** Eligibility criteria.

Inclusion criteria	Exclusion criteria
<ul style="list-style-type: none"> <li>• Age <math>\geq 18</math> years at diagnosis.</li> <li>• Diagnosis of rectal adenocarcinoma.</li> <li>• Radiologically measurable or clinically evaluable disease as defined by protocol.</li> <li>• ECOG Performance Status (PS): 0, 1 or 2.</li> <li>• For this patient, the standard treatment recommendation in the absence of a clinical trial would be combined modality neoadjuvant chemoradiation followed by curative intent surgical resection.</li> <li>• Candidate for sphincter-sparing surgical resection prior to initiation of neoadjuvant therapy according to the primary surgeon.</li> <li>• Clinical stage: T2N1, T3N0, T3N1. <ul style="list-style-type: none"> <li>• N2 disease is to be estimated as four or more lymph nodes that are <math>\geq 10</math> mm.</li> <li>• Clinical staging should be estimated based on the combination of the following assessments: physical exam by the primary surgeon, CT or PET/CT scan of the chest/abdomen/pelvis and either a pelvic MRI or an ultrasound (ERUS). If a pelvic MRI is performed, it is acceptable to perform CT of the chest/abdomen, omitting CT imaging of the pelvis.</li> </ul> </li> <li>• The following laboratory values obtained were <math>\leq 28</math> days prior to registration: <ul style="list-style-type: none"> <li>• Absolute neutrophil count (ANC) <math>\geq 1500/\text{mm}^3</math></li> <li>• Platelet count <math>\geq 100,000/\text{mm}^3</math></li> <li>• Hemoglobin <math>&gt; 8.0</math> g/dL</li> <li>• Total bilirubin <math>\leq 1.5 \times</math> upper limit of normal (ULN)</li> <li>• SGOT (AST) <math>\leq 3 \times</math> ULN</li> <li>• SGPT (ALT) <math>\leq 3 \times</math> ULN</li> <li>• Creatinine <math>\leq 1.5 \times</math> ULN</li> </ul> </li> <li>• Negative pregnancy test done <math>\leq 7</math> days prior to registration, for women of childbearing potential only.</li> <li>• Patient of childbearing potential is willing to employ adequate contraception. Appropriate methods of birth control include abstinence, oral contraceptives, implantable hormonal contraceptives or double barrier method (diaphragm plus condom).</li> <li>• Provide informed written consent.</li> <li>• Willing to return to enrolling medical site for all study assessments.</li> <li>• Primary rectal tumors that are within 5–12 cm of the anal verge (dropped in a protocol amendment).</li> </ul>	<ul style="list-style-type: none"> <li>• Clinical T4 tumors.</li> <li>• Primary surgeon indicates need for abdominoperineal resection (APR) at baseline.</li> <li>• Evidence that the tumor is adherent to or invading the mesorectal fascia on imaging studies such that the surgeon would not be able to perform an R0 resection (one with negative margins).</li> <li>• Tumor is causing symptomatic bowel obstruction (patients who have had a temporary diverting ostomy are eligible).</li> <li>• Chemotherapy within 5 years prior to registration. (Hormonal therapy is allowable if the disease-free interval is <math>\geq 5</math> years.)</li> <li>• Any prior pelvic radiation.</li> <li>• Other invasive malignancy <math>\leq 5</math> years prior to registration. Exceptions are colonic polyps, non-melanoma skin cancer, ductal carcinoma in situ, bladder carcinoma in situ or carcinoma in situ of the cervix.</li> <li>• Any of the following because this study involves an agent that has known genotoxic, mutagenic and teratogenic effects. <ul style="list-style-type: none"> <li>• Pregnant women</li> <li>• Nursing women</li> <li>• Men or women of childbearing potential who are unwilling to employ adequate contraception</li> </ul> </li> <li>• Co-morbid illnesses or other concurrent disease which, in the judgment of the clinician obtaining informed consent, would make the patient inappropriate for entry into this study or interfere significantly with the proper assessment of safety and toxicity of the prescribed regimens.</li> </ul>

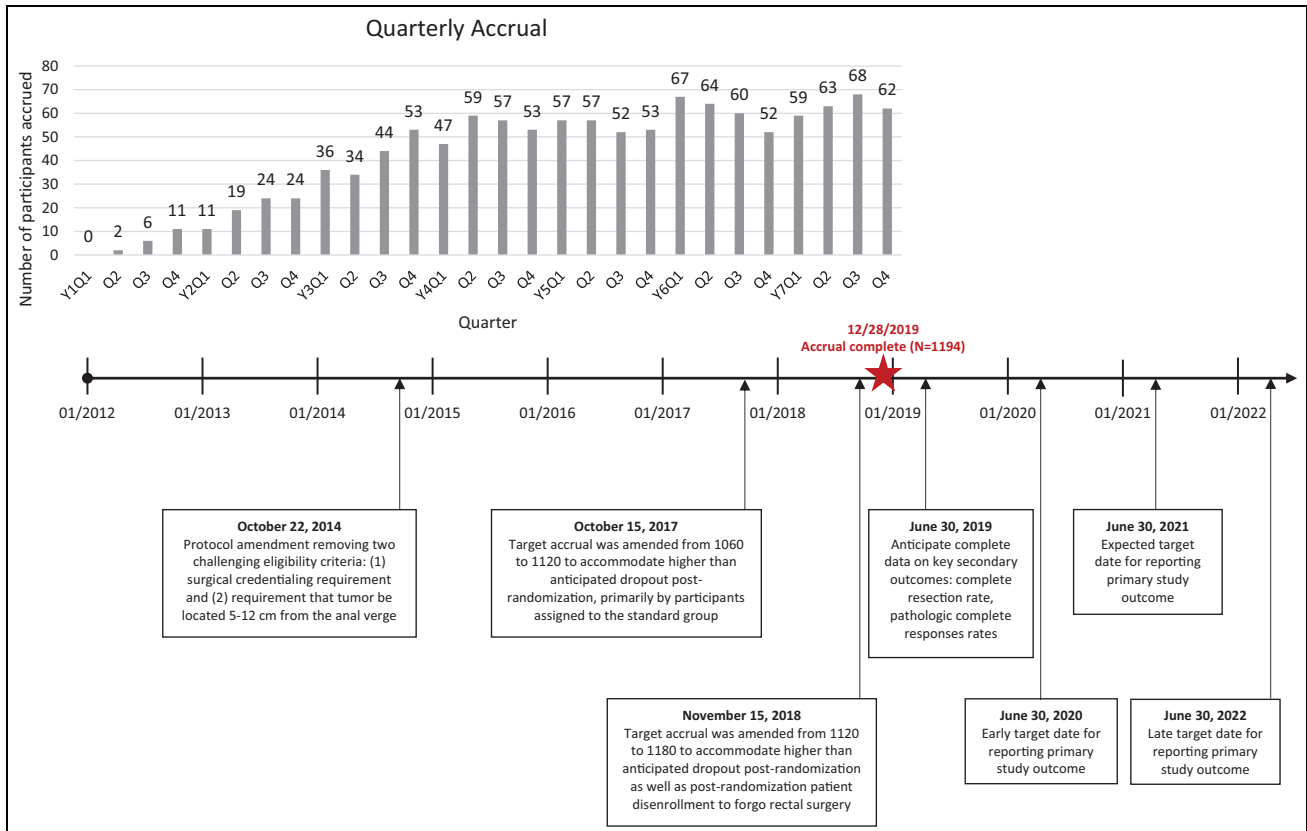
ECOG: Eastern Cooperative Oncology Group; CT: computed tomography; PET: positron emission tomography; SGOT: serum glutamic oxaloacetic transaminase; SGPT: serum glutamic pyruvic transaminase; AST: aspartate aminotransferase; ALT: alanine aminotransferase.

surgeons be “credentialed” in performing TME. Second, because community-based radiologists vary widely in their experience interpreting rectal magnetic resonance imaging (MRI), central radiology review was viewed as essential for ascertaining eligibility and response to neoadjuvant FOLFOX. The protocol required study staff at each performance site to upload MRI images for central review by an expert subspecialty radiologist who focuses on rectal MRI. For patients in the intervention group, the magnitude of response was evaluated on a repeat MRI by central radiology review to identify poor responders ( $< 20\%$ )

to FOLFOX and offer them another chance to receive preoperative chemoradiation.

### ***Challenge #3: balancing prioritization of two endpoints, time to local and distant recurrence***

The primary goal of cancer treatment is to maximize overall survival. However, in adjuvant colorectal cancer trials, 3-year disease-free survival (DFS) is an established surrogate endpoint that is well correlated with overall survival.<sup>10</sup> The main purpose of pelvic radiation is to prevent local recurrences that are a particularly



**Figure 2.** Study accrual and timeline.

devastating complication of rectal cancer. Choosing DFS as a single primary endpoint was not felt to place sufficient emphasis on the importance of pelvic recurrence. Similarly, local recurrence-free survival was unacceptable as a primary endpoint alone because it does not consider the reality that most patients succumb to rectal cancer because of distant tumor spread. To resolve this, a sequential decision strategy considering both time to local recurrence and DFS (including distant recurrence) as co-primary endpoints was adopted.

#### **Challenge #4: ensuring complete high-fidelity capture of patient's symptoms**

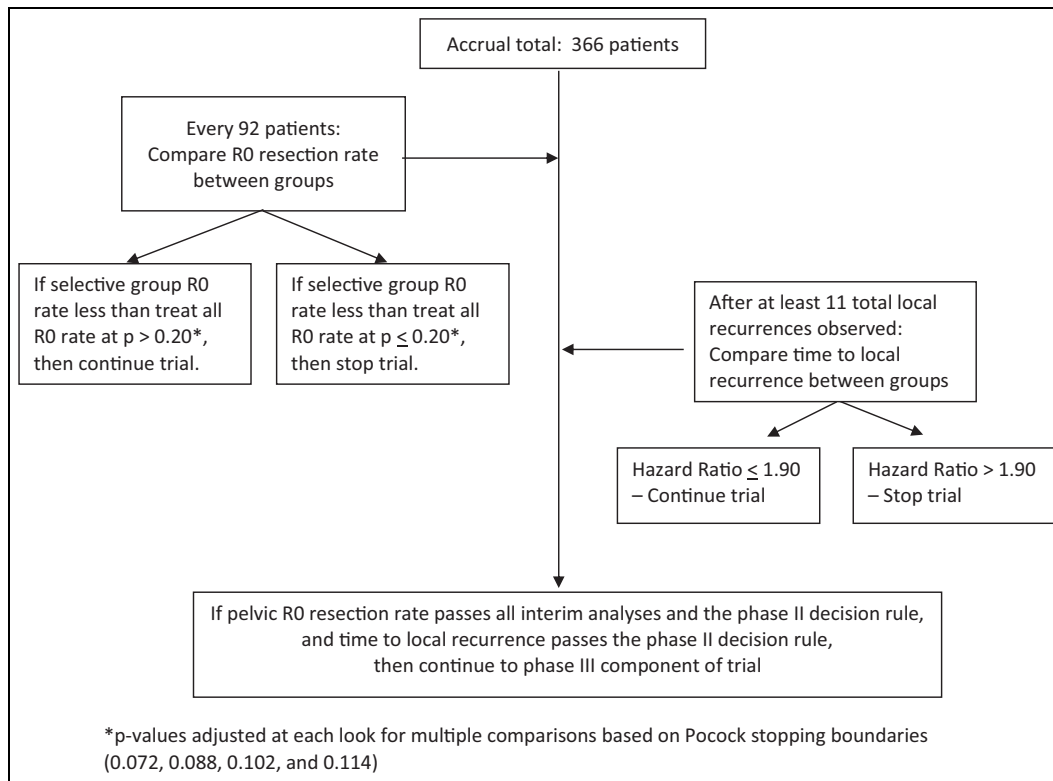
Chemoradiation, surgery and chemotherapy for rectal cancer each have substantial short- and long-term toxicities, but these are not consistently or well recorded by clinicians.<sup>11</sup> Because the study hypothesis was that neoadjuvant chemotherapy would have less toxicity than chemoradiation, reliable symptom reports were necessary. To accomplish this, participants enrolled at US and Canadian sites who were fluent in English or Spanish were asked to report 15 symptoms weekly during active treatment and every 6 months during follow-up using validated items from the NCI's Patient-Reported Outcomes Version of the Common

Terminology Criteria for Adverse Events (PRO-CTCAE).<sup>12</sup>

## **Results**

All participants in PROSPECT are randomized 1:1 to either neoadjuvant combined modality therapy with 5FU or capecitabine and concurrent pelvic radiation (standard treatment) or to six cycles of initial preoperative mFOLFOX6 with selected use of radiation depending on response to chemotherapy (intervention). Subject to passing all safety and initial efficacy stopping rules, accrual continued uninterrupted between the phase II and III portions of the trial and the participants in phase II portion will be included in phase III statistical testing and inferencing. The study hypothesis was that neoadjuvant FOLFOX does not compromise the rate of complete resection, local control, disease-free or overall survival in comparison with neoadjuvant chemoradiation.

PROSPECT was activated in January 2012, but few sites opened before 2013 due to reorganization of the NCI clinical trial networks, a new system for submitting study forms and the cumbersome surgeon credentialing. Figure 2 shows accrual and protocol amendments. Despite the challenges of randomization, the study has met its accrual goal by addressing four key challenges.



**Figure 3.** Phase II analytic approach.

### ***Approach to challenge #1: phase II/III hybrid design with stringent stopping criteria***

Because of oncologists' concerns about compromising cure rates for patients randomized to the intervention, a phase II/III design was selected with distinct outcomes for each phase. The primary outcome of the phase II component was to assure that the intervention group does not produce either inferior rate of complete resection (R0 resections predict disease-free and overall survival) or inferiority for time to local recurrence, compared to standard treatment. The R0 resection rate was defined as the number of patients with a complete resection divided by the total number in the analysis population. Time to local recurrence was defined as the time from randomization to the first date of local recurrence. If interim stopping criteria were not met, the phase II was to accrue 366 patients and proceed without interruption to phase III. The phase III primary outcome was a co-primary endpoint of the time to local recurrence-free survival and DFS and required 1060 patients. Target accrual was amended to 1120 on 15 October 2017 to accommodate higher than anticipated dropout post-randomization, primarily by participants assigned to the standard group who declined pelvic chemoradiation. Among the first 991 enrolled participants, 7.7% and 1.0% in standard treatment and intervention group, respectively, withdrew consent after randomization; 8% were deemed ineligible by central imaging review.

Phase II decision rules were motivated by safety concerns (see Figure 3). A sample size of 366 provided 82% power to detect whether the R0 resection rate (87%) in the intervention group was at least 6% (absolute difference) worse than that in the standard treatment group (93%), at the one-sided significance level of 0.20. After surgical outcomes were available on all 366 phase II participants, the study was to be terminated for a p value of  $\leq 0.114$ . Three interim analyses were performed when surgical outcomes became available on the first 92, 184 and 274 patients. The Pocock version of the Lan and DeMets stopping rules<sup>13–15</sup> allowed monitoring with a greater likelihood of early stopping for poor results in the intervention group. If the p value was less than or equal to 0.072, 0.088 and 0.102 at three interim looks, respectively, accrual was to be suspended. For time to local recurrence, the phase II decision algorithm was performed after 11 events were observed. If the hazard ratio (HR) for time to local recurrence was  $\leq 1.90$ , the study was to proceed to phase III. When all these criteria were met, PROSPECT proceeded seamlessly to phase III. All patients enrolled in phase II and III components will be used in the phase III analysis and decision-making. The Alliance Data and Safety Monitoring Board have not authorized communication of any results and closely track the study's progress. Results will be released when deemed appropriate.

## *Approach to challenge #2: simplification of unworkable eligibility criteria and central radiology review*

When PROSPECT was designed, colorectal cancer trialists worried that omission of chemoradiation would compromise outcomes for patients treated outside of specialty cancer centers. Accordingly, the study incorporated safeguards to ensure that high-risk patients would not be enrolled.

The first safeguard was the requirement that participating surgeons demonstrate technical proficiency at performing a TME by submitting photographs of up to 10 previous cases before they could be credentialed to operate on a study patient. The rationale for this requirement was that omission of chemoradiation is likely to be safe when high-quality surgical technique achieves clear margins. This safeguard was controversial from the outset and quickly proved unworkable as few colorectal surgeons had a cache of TME photos and were willing to submit them. The protocol was amended to drop surgeon credentialing (Figure 2) before 50 patients were accrued. However, to evaluate the quality of TME, surgeons were asked to submit photos of each participant's TME specimen, which were reviewed by two colorectal surgeons and evaluated for completeness. Because of this amendment which changed surgical credentialing to quality assurance, study results should have generalizability to rectal cancer patients treated in diverse settings.

The second safeguard was the requirement that tumors be located between 5 and 12 cm from the anal verge. This eligibility criterion was intended to exclude patients with distal tumors based on evidence that these patients are at higher risk for local recurrence. In contrast, patients with proximal tumors at the rectosigmoid junction are at minimal risk of local recurrence; therefore, the goal of the 5–12 cm tumor location was to obtain a homogeneous group of rectal tumors—neither too distal nor too proximal. Although workable in the phase II single-center pilot, this criterion proved unworkable in the multicenter setting. A single patient's tumor location was interpreted differently by gastroenterology, imaging, surgery, radiation and medical oncology, and research staff were challenged to discern which interpretation was the truth. Moreover, variation in anatomy and patient body mass made this criterion unworkable. It was dropped in an amendment (Figure 2) in favor of an alternative that proved simpler to operationalize; the surgeon had to indicate that the patient (1) was a candidate for a low anterior resection with a sphincter preserving TME and (2) would receive chemoradiation in the absence of a clinical trial. This excluded patients with very distal tumors that would require an abdominoperineal resection and proximal tumors that might not need radiation at all; this proved feasible for sites to implement with fidelity.

A third safeguard was exclusion of patients with clinical N2 tumors. When the study was designed, some rectal cancer experts felt strongly that patients with clinical N2 tumors should receive standard chemoradiation. Conversely, others felt that patients with clinical N2 tumors were more likely to benefit from a neoadjuvant chemotherapy approach. Reflecting the reality that clinical trial design involves political compromise, patients with clinical N2 tumors were excluded. However, ascertaining clinical N2 tumors (four or more lymph nodes with tumor) is subject to variability in judgment calls about what constitutes a positive node on imaging. To achieve consistency across sites, investigators defined clinical N2 tumors as those with four or more regional lymph nodes each  $\geq 10$  mm on pelvic MRI.

The fourth safeguard was review of baseline images by a central radiologist expert in interpretation of rectal MRI. Each participant's staging MRI or computed tomography (CT) (CT and endorectal ultrasound was allowed for patients unable to undergo MRI) was electronically submitted to the Alliance imaging core. Local radiologists in conjunction with site investigators were responsible for primary determination of eligibility and final decisions. However, the central radiologist re-interpreted each image blinded to treatment assignment. If the central reviewer regarded the patient as ineligible, notification was sent to the primary study investigator and accruing physician describing the discrepancy (e.g., suspected T4 tumor, involved radial margin or more than four lymph nodes  $\geq 10$  mm). Sites then made a final decision about whether the patient was eligible. The central radiology reviewer also re-reviewed all restaging studies for patients assigned to the intervention arm and categorized the magnitude of response as  $\geq 20\%$  or  $< 20\%$ . Ideally, images would be submitted before study treatment initiation and the central review would take place immediately, but to maximize the ease of participation, strict time windows were not required and thus review was not performed in "real time." Although the majority of baseline central imaging reviews were performed within 2 weeks of study enrollment, primary responsibility for eligibility determination was left to the sites. The central radiologist regarded approximately 8% of participants as ineligible but more frequently disagreed about scan interpretation (such as the involvement of lymph nodes) that would not affect eligibility. The study has amassed an annotated imaging bank with case report forms interpreted by local radiologists and a specialty radiologist that will provide insight into the radiologic markers of treatment outcomes. In all primary analyses, clinical staging and response criteria are determined by site review with advisory input from central review. Secondary analyses will examine (1) the site independent review (before advice from the central reviewer) and (2) the central review.

### *Approach to challenge #3: designing a decision algorithm that jointly considers two co-primary endpoints*

To balance the importance of time to local recurrence-free survival and DFS in rectal cancer, co-primary endpoints with a sequential decision algorithm were chosen for the phase III analysis. The intervention strategy will be favored if it either

1. Achieves superior DFS compared to the standard group (regardless of the time to local recurrence results) or
2. Is at least non-inferior to the standard group for both DFS and time to local recurrence.

The phase III sequential hypothesis-testing procedure for the phase III component will first compare DFS between the two groups for non-inferiority of the intervention group. If non-inferiority is supported, then the intervention will be tested for superiority for DFS based on a one-sided test. If non-inferiority is not demonstrated, standard treatment will be declared preferred. If superiority of the intervention group for the DFS endpoint is determined, further formal testing will stop for the primary aims, and we will conclude that the intervention strategy is preferred. However, if non-inferiority but not formal superiority of the intervention is found based on decision rules for the DFS endpoint, then the co-primary endpoint of time to local recurrence will be tested for non-inferiority. If non-inferiority in time to local recurrence is supported, the intervention will be declared preferred. Otherwise, standard treatment will be declared preferred. This decision algorithm is shown in Figure 4, and the expected rates that form the basis for the power calculations are shown in Table 2. In total, 1000 patients provide 85% power to detect non-inferiority of the DFS and time to local recurrence jointly at the overall alpha level of 0.05, if the true DFS in the selective use group is slightly superior (approximately 2% absolute percentage superior at 3 years) to the treat-all approach (equivalent to  $HR = 0.91$ ) and the true local recurrence-free survival in the selective use group is the same as in the treat-all group (equivalent to  $HR = 1$ ). The final analysis will be conducted when there are at least 406 and 75 events observed for DFS and time to local recurrence, respectively. The final sequential decision rules are listed as follows:

1. If the HR comparing DFS in the intervention group to the standard group is greater than 1.115 (in favor of the standard group), then the standard group is declared to be preferred; otherwise, the analysis proceeds to the following steps.
2. If the HR comparing DFS in the intervention group to the standard group is  $<0.8367$ , then the

intervention is declared to be preferred; otherwise, it proceeds to the following step.

3. If the HR comparing time to local recurrence in the intervention to the standard group is  $\leq 1.44$ , then the intervention is preferred; otherwise, the standard is preferred.

Three interim analyses testing for superiority in DFS (selective use versus treat-all, two-sided log-rank test) will be conducted when 102, 230 and 305 events have been observed. The O'Brien-Fleming version of the Lan and DeMets stopping rules will be implemented for the DFS interim analyses.<sup>16</sup> If the p value of the two-sided log-rank test is  $\leq 0.001$ , 0.0077, 0.0246 at the three interim looks, respectively, the study will be stopped early and will conclude that the intervention group is preferred if the HR is less than 1, or the standard group is preferred if the HR is greater than 1. The sample size estimation was revised to require a total of 1180 patients to be randomized. Power calculations require 500 participants in each group and thus allow for up to 180 cancellations and major eligibility violations. All analyses on secondary endpoints will be conducted on per-protocol population, except safety endpoints. The intention-to-treat population will be used for sensitivity analyses.

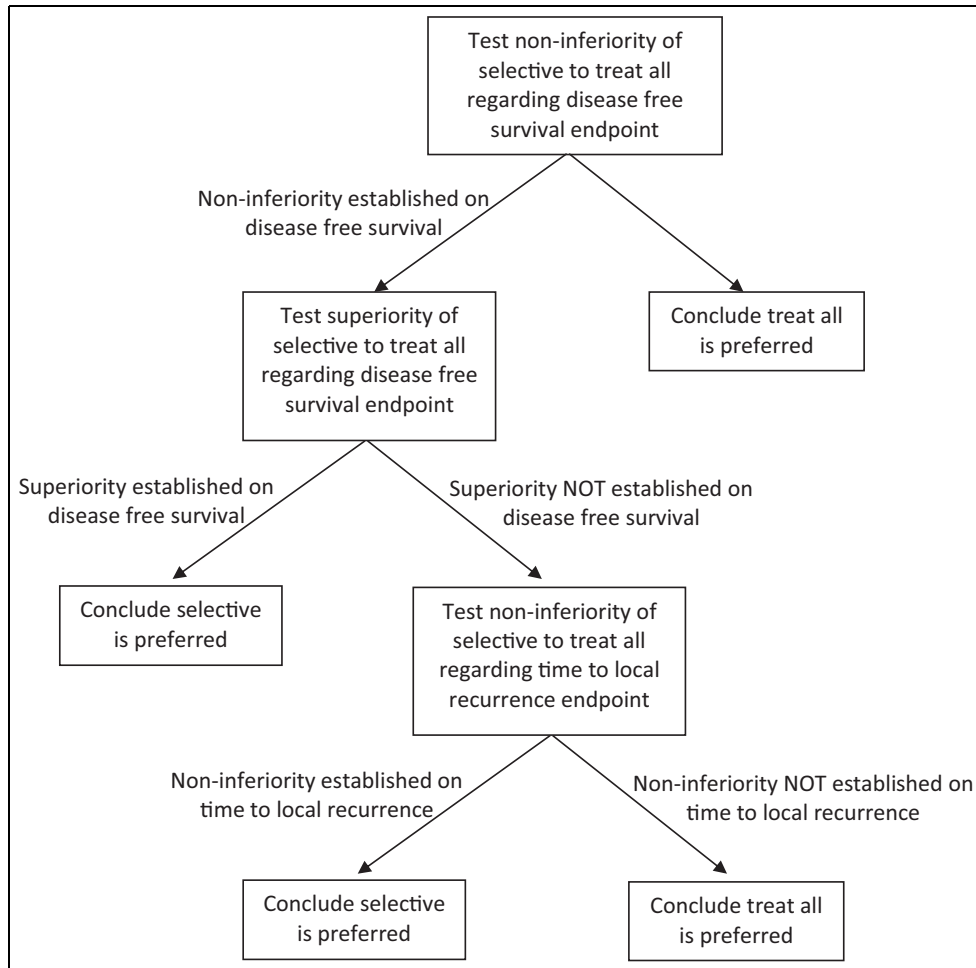
### *Approach to challenge #4: prompting patients to report symptom events by phone or email*

The PROSPECT protocol mandated expedited reporting of serious adverse events. Routine adverse event reporting was collected at baseline and at each evaluation of dysphagia, diarrhea, constipation, nausea, vomiting, oral mucositis, dyspnea, peripheral sensory neuropathy, pain, fatigue, anorexia, anxiety and depression using the CTCAE version 4.0. Based on evidence that patient and clinician reports of symptom burden differ,<sup>17</sup> PROSPECT also captured patients' direct toxicity reports of these same symptoms using PRO-CTCAE. Participants were given the choice by centralized PRO-CTCAE staff to use a web-based questionnaire or Interactive Voice Response System by telephone to report their symptoms once at baseline, weekly during active preoperative treatment and once every 6 months for 3 years following surgery. Patients receive up to two reminders and non-respondents then receive a phone call from a centralized coordinator. During active treatment, the response rate is  $>77\%$  without and rises to  $>92\%$  with the phone coordinator backup calls. In addition, sites administer a quality of life questionnaire to each participant at baseline, 1–2 weeks prior to surgery and 12 and 24 months after rectal resection.

## **Discussion**

The PROSPECT investigators' experience with both successful and unsuccessful innovations may benefit





**Figure 4.** Phase III analytic approach.

**Table 2.** Statistical hypotheses for the Phase III PROSPECT trial (N1048).

	Disease-free survival	Time to local recurrence
Hazard ratio comparing intervention to standard treatment groups	$\geq 1.23$	$\geq 1.775$
Based on design assumption corresponds to	A 3-year DFS rate in the intervention group of 69% versus 74% in the standard group	A 3-year local recurrence-free rate of 93% in intervention versus 96% in the standard group

DFS: disease-free survival.

other clinical trialists. First, PROSPECT used a seamless phase II/III design, an alternative to sequential phase II and III trials that is appropriate when (1) a rapidly obtained phase II endpoint correlated with the primary phase III endpoint is available, (2) positive phase II results would provide sufficient motivation to launch the phase III component and (3) phase III resources and infrastructure are in place.<sup>18</sup> Although impossible to know how other designs would have worked, a multi-center single-arm phase II would have been critiqued

for “selection,” and sequential phase II and III studies would have required more patients and time.

Second, PROSPECT’s overly complex eligibility criteria were not a sufficient antidote for meager phase II data. PROSPECT initially included requirements such as surgeon credentialing in TME and precise anatomic location of the primary rectal tumor based on compromise among surgeons, radiation and medical oncologists. These criteria proved unworkable and were removed in an early protocol amendment. A key lesson

learned in PROSPECT is that multidisciplinary care varies in how it is organized. In some centers, rectal cancer patients see surgeons first; in others, it is radiation or medical oncology. Initially, there was some confusion among clinical research professionals about how to resolve ambiguities and discrepant interpretations among team members about response and eligibility criteria. Clarification that the primary surgeon adjudicated whether a patient was eligible for TME streamlined study conduct. Uploading films for review by a central radiologist was straightforward for sites to implement and has been performed with high adherence. Although PROSPECT did not perform rapid real-time central review prior to randomization, the interactive, monitored and scheduled reading sessions between the core lab and the central radiologist allowed for a seamless feedback loop between central and site-specific teams. Moreover, the study image bank provides a mechanism for quality control superior to the prevailing method that relies on imaging reports. It also provides a repository for imaging biomarker discovery. The number of discrepant interpretations and whether disagreements pertain to tumor size, lymph node involvement or margin status will inform decisions about incorporating central imaging review into future rectal cancer trials and may have implications for radiology training in pelvic MRI.

Third, the statistical design that simultaneously weighs two endpoints of interest, DFS and time to local recurrence is a pragmatic approach that attends to the competing goals of rectal cancer management. Although sequential hypothesis testing and a decision tree is more complex to explain than the more familiar approach reliant on a primary and then a secondary endpoint, this strategy is more relevant to the clinical decision-making process and trade-offs faced by clinicians.

Fourth, PROSPECT is the first cooperative group randomized trial to include systematic toxicity and adverse event reporting directly by patients. Direct patient reporting by phone or web proves to be a feasible and efficient approach to obtaining symptom reports. Factors that contribute to the success of PRO-CTCAE implementation include keeping the number of symptoms elicited short, weekly requests, having telephone backup for missed assessments and providing participants with a choice of participation mode (phone or web). Central coordination of PRO-CTCAE means that the burden on sites is minimal. However, patients' adverse event reports are not systematically reviewed in real time by clinical teams and therefore cannot be used to ameliorate symptom burden.

Finally, although designed as a traditional explanatory study, some aspects of the study design resemble a pragmatic trial. Recognizing regional variation in rectal cancer delivery, the study allowed flexibility in several areas not expected to influence study outcomes. For example, pelvic MRI was preferred for staging, but CT

with endorectal ultrasound was allowed in recognition of the fact that some patients have large body mass or hardware that precludes MRI. Medical oncologists could use either oral capecitabine or intravenous 5FU for chemosensitization during radiation. Radiation oncologists could choose standard or intensity-modulated radiation therapy and surgeons could use open or laparoscopic approaches. Although six cycles of FOLFOX in the intervention arm and eight cycles in the standard group were suggested, the specific adjuvant regimen and duration was discretionary and dose modification guidelines were flexible. PROSPECT's explanatory features are intended to yield a rigorous and compelling comparison of treatment alternatives, and its pragmatic features should maximize generalizability to a wide variety of oncology practice settings. In summary, obstacles in the design and conduct of PROSPECT and the strategies used to overcome them may inform trialists evaluating other complex treatment interventions, particularly trials where implementation of consistent criteria for eligibility and outcomes across hundreds of practice settings is necessary.

### **Acknowledgements**

Dr. Daniel Sargent was integral to the statistical design of PROSPECT and although his untimely death did not allow him to participate in preparation of this manuscript, his posthumous contributions to this clinical trial are hereby recognized. Without the efforts of Deb Papenfus, protocol quality assurance specialist at the Mayo Clinic, study data would not be high quality or complete. Without the efforts of Jennifer Wind, program manager at Dana-Farber Cancer Institute, study accrual and execution would not have been accomplished. Members of the Gastrointestinal Cancer Committee of the Alliance for Clinical Trials in Oncology and of the NCI's Colorectal Cancer Task Force provided critical input and feedback. Finally, efforts from hundreds of clinicians from multiple disciplines and sites, dedicated clinical research professionals and most importantly patient participants have made the conduct and execution of PROSPECT possible.

### **Declaration of conflicting interests**

The author(s) declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

### **Funding**

Research reported in this publication was supported by the National Cancer Institute of the National Institutes of Health under award numbers U10CA180821 and U10CA180882 (to the Alliance for Clinical Trials in Oncology), U10CA180791, U10CA180838 and U10CA180867. The PRO-CTCAE component is funded under an existing NCI contract (HHS-N261201000063C; PI: E.B.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

## Trial registration

ClinicalTrials.gov Identifier: NCT01515787. Trial sponsor: National Cancer Institute (NCI)'s Clinical Therapy Evaluation Program (CTEP), The Alliance for Clinical Trials in Oncology.

## References

1. NIH and consensus conference. Adjuvant therapy for patients with colon and rectal cancer. *JAMA* 1990; 264: 1444–1450.
2. NQF-Endorsed Measures (QPS). National Quality Forum.
3. Sauer R, Becker H, Hohenberger W, et al. Preoperative versus postoperative chemoradiotherapy for rectal cancer. *N Engl J Med* 2004; 351(17): 1731–1740.
4. Carne PW and Nelson H. Are we overtreating rectal cancer: time for another trial. *Ann Surg Oncol* 2004; 11(2): 124–126.
5. Carlsen E, Schlichting E, Guldvog I, et al. Effect of the introduction of total mesorectal excision for the treatment of rectal cancer. *Br J Surg* 1998; 85(4): 526–529.
6. Enker WE, Merchant N, Cohen AM, et al. Safety and efficacy of low anterior resection for rectal cancer: 681 consecutive cases from a specialty service. *Ann Surg* 1999; 230(4): 544–552, discussion, 552–554.
7. MacFarlane JK, Ryall RD and Heald RJ. Mesorectal excision for rectal cancer. *Lancet* 1993; 341: 457–460.
8. Merchant NB, Guillem JG, Paty PB, et al. T3N0 rectal cancer: results following sharp mesorectal excision and no adjuvant therapy. *J Gastrointest Surg* 1999; 3(6): 642–647.
9. Schrag D, Weiser MR, Goodman KA, et al. Neoadjuvant chemotherapy without routine use of radiation therapy for patients with locally advanced rectal cancer: a pilot trial. *J Clin Oncol* 2014; 32(6): 513–518.
10. Sargent DJ, Patiyil S, Yothers G, et al. End points for colon cancer adjuvant trials: observations and recommendations based on individual patient data from 20,898 patients enrolled onto 18 randomized trials from the ACCENT Group. *J Clin Oncol* 2007; 25(29): 4569–4574.
11. Basch E, Jia X, Heller G, et al. Adverse symptom event reporting by patients vs clinicians: relationships with clinical outcomes. *J Natl Cancer Inst* 2009; 101(23): 1624–1632.
12. Basch E, Reeve BB, Mitchell SA, et al. Development of the National Cancer Institute's patient-reported outcomes version of the common terminology criteria for adverse events (PRO-CTCAE). *J Natl Cancer Inst* 2014; 106(9): dju244.
13. Pocock SJ. Group sequential methods in the design and analysis of clinical trials. *Biometrika* 1977; 64: 191–199.
14. Lan KKG and DeMets DL. Discrete sequential boundaries for clinical trials. *Biometrika* 1983; 70: 659–663.
15. Pocock SJ. Interim analyses for randomized clinical trials: the group sequential approach. *Biometrics* 1982; 38(1): 153–162.
16. O'Brien PC and Fleming TR. A multiple testing procedure for clinical trials. *Biometrics* 1979; 35(3): 549–556.
17. Falchook AD, Green R, Knowles ME, et al. Comparison of patient- and practitioner-reported toxic effects associated with chemoradiotherapy for head and neck cancer. *JAMA Otolaryngol Head Neck Surg* 2016; 142(6): 517–523.
18. Wang M, Dignam JJ, Zhang QE, et al. Integrated phase II/III clinical trials in oncology: a case study. *Clin Trials* 2012; 9(6): 741–747.