Outcomes and endpoints in trials of cancer treatment: the past, present, and future

Michelle K Wilson, Katherine Karakasis, Amit M Oza

Cancer treatment should allow patients to live better or longer lives, and ideally, both. Trial endpoints should show clinically meaningful improvements in patient survival or quality of life. Alternative endpoints such as progression-free survival, disease-free survival, and objective response rate have been used to identify benefit earlier, but their true validity as surrogate endpoints is controversial. In this Review we discuss the measurement, assessment, and benefits and limitations of trial endpoints in use for cancer treatment. Many stakeholders are affected, including regulatory agencies, industry partners, clinicians, and most importantly, patients. In an accompanying Review, reflections from individual stakeholders are incorporated into a discussion of what the future holds for clinical trial endpoints and design.

Lancet Oncol 2015; 16: e32–42

Princess Margaret Cancer Centre, Toronto, Canada (M K Wilson FRACP, K Karakasis MSc, Prof A M Oza FRCP)

Correspondence to: Prof Amit M Oza, University of Toronto Princess Margaret Cancer Centre, Toronto, M5G 2M9, Canada
amit.oza@uhn.ca

Introduction

Cancer treatment aims to enable patients to live longer and better lives than they would without treatment.[1] Accordingly, the gold standard to assess therapeutic efficacy is a statistically signifant and clinically meaningful improvement in overall survival, quality of life, or decrease in cancer-related symptoms. The issue is who defines what is clinically meaningful, with substantial variation in opinions between patients, clinicians, and regulatory agencies.[1]

Overall survival is an unambiguous outcome measure. However, as an unequivocal marker of therapeutic benefit, it is controversial. Improvement in the efficacy of cytotoxic and novel treatments, and a better understanding of disease biology, means that some cancers now behave like chronic diseases. The detection of differences in overall survival in trials can be complex with confounding results related to trial duration, cost, sample size, and treatment choices after progression.[2] If cancer is regarded as a chronic disease, new treatments need to have clinically meaningful outcomes for patients, such as improved quality of life, symptom control, and tolerability of treatment.

Endpoints such as progression-free survival (PFS) and time to progression are used because they provide more rapid measures of difference in treatment effect than overall survival. A robust debate continues on whether these are meaningful outcomes in their own right, or purely surrogate endpoints for overall survival and disease control. Unlike overall survival, endpoints such as PFS and time to progression are less affected by subsequent treatments, palliative care, and comorbidities.[2] By contrast with death, disease progression as an endpoint is continuous, but is identified and measured by discrete clinical or radiological assessments, and hence is dependent on the timing of these investigations. In routine practice, disease status is established by clinical, biochemical, and radiological factors; conversely, trial outcomes are often defined by radiological measures alone to increase objectivity.

Cancer is a multitude of diseases with different natural histories and clinical characteristics.[3,4] Therefore, trial design needs to accommodate the inherent properties of the disease and patient population to ensure meaningful clinical outcomes. This approach will help with the identification of smaller homogenous subpopulations that benefit from a treatment.[3,5] We review present practices in the measurement of clinical outcomes in oncology trials and assess these practices in the context of modern clinical oncology. This Review also addresses the objectivity, reliability, and validity of the methods used to assess endpoints in clinical trials and the importance of integration of these techniques into the study design.

Endpoints in clinical trials

Views on the importance of overall survival and PFS as outcome measures in trials are polarised. Treatment effectiveness has been defined as a clinically meaningful benefit to the patient with the objective of the patient living for longer or better, or both, than if they did not receive the treatment.[1] This measure of efficacy might be shown by symptomatic improvement, or improvement in established disease progression endpoints. The characteristics of the disease and treatment (prognosis, aggressiveness of disease, symptoms, and salvage therapy) need to be applied to trial design to define appropriate endpoints. For example, in metastatic gall bladder cancer—with few therapeutic options and a median overall survival of less than 1 year, even a modest gain in overall survival might be clinically relevant.[6] In an indolent disease such as ovarian granulosa-cell tumour, with a median overall survival of more than 15 years and a tendency to late relapse,[7] the detection of therapeutic benefit in terms of overall survival is not realistic and a short term, more clinically meaningful objective measure of benefit might instead be improvement in symptoms or quality of life.

Overall survival and surrogate endpoints

Overall survival is a precise endpoint that shows patients alive after a fixed duration of time (table).[8,9,11] Effects from crossover, supportive care, and treatment after progression make measurement of overall survival increasingly protracted.[2] To measure more modest differences in overall survival, phase 3 trials have become larger, and more

resource intensive; present trial designs are challenging this approach.

Surrogate endpoints (table) have garnered interest as indirect measures of clinical benefit because they can generally be achieved in a shorter timeframe than overall survival; however, they carry the risk of misleading conclusions due to erroneous extrapolation from the results.[12] The definition of a surrogate endpoint is complex and controversial. According to the NIH Biomarkers Definitions Working Group, a clinical endpoint is a characteristic or variable that reflects how a patient feels, functions, or survives, and a surrogate endpoint is defined as a biomarker intended to act as a clinical endpoint.[13]

Controversy also exists as to whether these surrogate endpoints can be regarded as clinical endpoints in their own right, or only as surrogate endpoints. The Prentice criteria provide guidelines[14] for the validation of surrogate endpoints, and serve to describe the relation between the intervention and endpoints of interest (both surrogate and clinical). A key criterion in this validation is the assumption that the effect of the intervention on the surrogate endpoint also results in meaningful changes for the clinical endpoint.[14]

When designing a clinical trial, the decision to use a surrogate endpoint must consider both the advantages of a shorter trial duration (drawing conclusions sooner than it would be possible with established clinical endpoints) and the possibility of increased uncertainty from extrapolating how a surrogate endpoint would behave from historic trial data.[12,15] If the uncertainty outweighs the advantages of a shorter trial, then it would not be useful to use a surrogate endpoint as this might not reflect true clinical outcomes for these patients. However, if there is evidence that a surrogate endpoint would be of benefit, other factors should still be considered, such as would the biological mechanisms by which the treatment affects the surrogate endpoints and the clinical endpoints be similar in the new and previous trials? Also, will patient management after the surrogate endpoint is reached be similar to previous trials? Finally, will serious or detrimental side-effects arise in the time of observation until the surrogate endpoint is reached compared with the time the clinical endpoint is reached?[12] In practice, uncertainty remains in addressing these questions. For example, often little is known about long-term toxic effects with novel treatments.[16] A review of the updated labels of targeted therapies by the US Food and Drug Administration (FDA) showed that 39% of

	Regulatory approval	Study design	Advantages	Disadvantages
Overall survival	Regular approval (clinical benefit)	Randomised studies; masking not essential	Universally accepted measure of clinical benefit; easily measured; precisely measured	Can include large patient numbers; can be affected by crossover therapy and sequential therapy; includes non-cancer deaths
Symptom endpoints (patient-reported outcomes)	Regular approval (clinical benefit)	Randomised, masked studies	Patient perspective of clinical benefit	Masking is often difficult; data are often missing or incomplete; clinical relevance of small changes is unknown; multiple analyses; lack of validated instruments; often reported as mean or median group scores rather than individual results
Disease-free survival	Accelerated approval (surrogate endpoint)*	Randomised studies; used in adjuvant setting	Masked review recommended; smaller sample size and shorter follow-up than overall survival	Not statistically validated as a surrogate endpoint for survival in all settings; not precisely measured; subject to bias, especially in open-label studies; definitions vary between studies
Progression-free survival (includes all deaths); time to progression (deaths before progression excluded); progression-free survival 2 (includes all deaths); or time to second progression (deaths before progression excluded)	Accelerated approval or regular approval (surrogate endpoint)*	Randomised studies; masking preferred in comparative studies; masked review recommended; used in advanced setting	Smaller sample size and shorter follow-up than overall survival; includes measurement of stable disease; not affected by crossover or subsequent treatments (progression-free survival 2 less affected than overall survival); generally based on objective and quantitative assessment	Not statistically validated as a surrogate endpoint for survival in most settings; not validated as measure of quality of life; not precisely measured; subject to assessment bias, especially in open-label studies; definitions vary between studies; frequent radiological or other assessments; timing of assessments in treatment groups needs to be balanced; affected by censoring of data
Objective response rate	Accelerated approval or regular approval (surrogate endpoint)*	Single-arm or randomised studies; masking preferred in comparative studies; masked review recommended	Can be assessed in single-arm studies; assessed earlier and in smaller studies than overall survival; effect attributable to drug and not natural history of disease; used in advanced setting	Only a subset of patients benefit; not a direct measure of clinical benefit; not a comprehensive measure of drug activity; does not measure duration of clinical benefit
Complete response	Accelerated approval or regular approval (surrogate endpoint)*	Single arm or randomised studies; masking preferred in comparative studies; masked review recommended	Can be assessed in single-arm studies; durable complete responses can represent clinical benefit; assessed earlier and in smaller studies than overall survival	Not a direct measure of clinical benefit; not a comprehensive measure of drug activity; small subset of patients benefit
Clinical benefit rate	Not generally used in regulatory approval	Single-arm or randomised studies; masking preferred in comparative studies; masked review recommended	Can be assessed in single-arm studies; assessed earlier and in smaller studies than overall survival; includes complete response, partial response, and stable disease for a defined period	Not a direct measure of clinical benefit; not a comprehensive measure of drug activity; definition of duration of stable disease varies between studies; stable disease can reflect inherent characteristics of tumour rather than disease activity

Adapted from Pazdur,[8] McKee and colleagues,[9] and the CHMP Scientific Advisory Group for Oncology.[10] *Adequacy as a surrogate endpoint for accelerated approval or regular approval is very dependent on other factors such as effect size, effect duration, and benefits of other available treatment.

Table: **Endpoints for approval of cancer therapies**

serious adverse drug reactions and 39% of potentially fatal adverse drug reactions were not described in randomised trials and 49% of serious adverse drug reactions and 58% of potentially fatal adverse drug reactions were not described on initial drug labels.[16]

Surrogate endpoints such as PFS (figure 1) are used to implicitly assess overall survival, but the relation between surrogate endpoints as a measure of clinical benefit such as control or palliation of symptoms in patients needs to be considered.[18,19] PFS is the most commonly used surrogate endpoint, defined as time from randomisation to tumour progression or death, whereas time to progression is defined as the time from randomisation to the time of disease progression and does not include death.[20] The exact timing of the progression is established by a discrete clinical or radiological assessment and also depends on the growth rate of a cancer. Other surrogate endpoints include disease-free survival or relapse-free survival in trials of adjuvant therapy, objective response rate, and time to progression.[8] New targeted drug therapies allow an intrapatient time to progression ratio that can be used to compare response to new interventions with responses to previous treatments. These measures are subject to measurement error, and assessment and attrition bias.[17,18,20]

PFS is an imperfect outcome measure but is central to the design of clinical trials. For some types of cancer, improvement in PFS might not show a true benefit to the patient, because surrogacy of PFS for overall survival has been shown infrequently and might depend on the context. A clinical benefit in terms of PFS might not translate into a clinical benefit in overall survival for several reasons. First, subsequent treatments probably dilute initial benefit.[21] Second, a change in tumour burden with defined disease progression might be insufficient to affect the time to death in some cancers.[18] Third, prolonged exposure to targeted therapies might lead to the evolution of tumours with a different phenotype, thus, offsetting any initial advantage from the treatment shown initially as PFS.[18]

Overall survival is comprised of PFS and survival after progression.[21] Potential interventions available at disease progression include crossover into a different group of the trial, treatment with an alternative drug, continuation with the same drug (if it is of symptomatic benefit), or no further treatment.[21] The heterogeneity of these options makes it difficult to assess the effect (if any) of the initial intervention on overall survival, due to confounding factors from each subsequent intervention.[21] Analyses and statistical models to assess the concordance between PFS and overall survival show that when median survival after progression is short (ie, <12 months), PFS seems a reasonable alternative endpoint to overall survival.[21] A strong correlation has been shown between PFS and overall survival in advanced colorectal and extensive-stage small-cell lung cancers.[22–25] By contrast, in diseases such as metastatic breast cancer or recurrent ovarian cancers with a long period of survival after progression, a divergent relation exists between PFS and overall survival because of the number of effective therapeutic options available after progression.[21,25] However, evidence to the contrary exists with clear survival benefit shown with the use of trastuzumab either alone[26] or in combination with pertuzumab[27] in HER2-positive metastatic breast cancer.

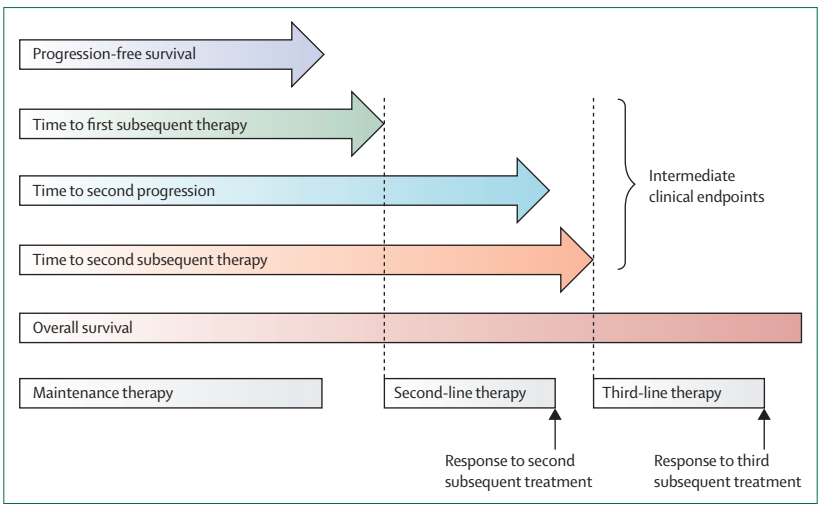

Figure 1: **Clinical endpoints used in cancer trials**
Adapted from Matulonis and colleagues[17] with permission from the American Cancer Society.

PFS is not affected by treatment crossover and subsequent treatments and therefore potentially allows a direct assessment of the effect of the intervention on disease control. Access to all potential therapeutic options that allow patients to live longer or better (or both) is desirable from a societal point of view. However, if the drug is investigated in a small and select population of patients and shows the same survival advantage for patients as in a large clinical trial, this strategy might then be more appropriate and potentially cost effective.

Good supportive care might also affect survival outcomes. Findings of a phase 3 study showed that the early introduction of palliative care for patients with metastatic non-small-cell lung cancer improved their quality of life and overall survival (8·9 months *vs* 11·6 months; p=0·02) even when age, performance status, and sex were taken into account (hazard ratio [HR] 1·7; p=0·01).[28] Although restricted to lung cancer, this study showed another variable that might affect patient survival after progression beyond the studied treatment period.

Trial endpoints can be prone to error and bias because they are contingent on consistent timing of tumour assessment in both control and intervention groups. Additionally, exclusion of patients from the analyses who were lost to follow-up or who started a new treatment can perpetuate bias.[29] An imbalance in the censoring between treatment groups and control groups will affect the outcome. For example, results of the BOLERO-2 study[30] showed that 24% of patients in the treatment group compared with 6% in the control group discontinued treatment; therefore, determination of the true efficacy in

terms of primary outcome (PFS) was difficult. Most statistical models assume the expected survival for censored patients is the same as that of patients who were not censored.[29]

Maintenance therapy in patients with no active disease or disease-related symptoms can make it a challenge to define and detect clinical benefit. In ovarian cancer, the use of PARP inhibitors shows the difficulty of designing trials with appropriate endpoints to show (or exclude) improved outcomes. Trials of PARP inhibitors have shown clinically convincing efficacy of monotherapy, particularly in women with germ-line *BRCA* mutations.[31–33] Durable benefit has been seen when these drugs were given continuously in the maintenance setting.[31,32] However, these patients are known to generally have disease that responds well to chemotherapy, and might have as many as ten different lines of treatment in a median 5–7 years. The design of a trial to show overall survival benefit in this scenario can be a challenge.[31]

So what is the optimum endpoint for trials of treatments that include the following characteristics: rely on chronic maintenance therapy; might have modest side-effects and impair quality of life in the long term, but to a lesser extent than short-term chemotherapy; and for which survival benefit to patients is hard to show because of effective (and alternative) therapies after progression? PARP inhibitors seem to have clinically meaningful benefit, but defining and measuring this benefit objectively remains difficult. Delayed disease progression might decrease a patient's emotional distress but needs to be balanced against drug toxicity and patient preference. This situation emphasises the difficulty of showing what treatment allows a patient to live with an overall better quality of life and symptom control. Overall survival might not be the best endpoint in the setting of maintenance therapy, which shows the importance of linking the question to the choice of trial endpoints.

Other endpoints

Time to disease progression or treatment

Times to second or third progression have been used as alternative endpoints for overall survival in several trials (figure 1), and are recognised by the European Medicines Agency[10] as a viable endpoint for registered clinical drug trials. Time to second progression is defined as time from randomisation to the second objective disease progression (or death from any cause), and time to third progression is defined as time from randomisation to the third objective disease progression (or death from any cause). The preservation of clinical benefit to the patient after progression provides additional confidence that the observed benefit is maintained and does not negatively affect subsequent response to treatment. Time to second subsequent treatment or death, which is not reliant on radiological assessment, is also under investigation as a surrogate endpoint.[17,31] Time to second subsequent treatment or death might mirror clinical practice more closely than time to second progression (with the assumption that subsequent treatment is started because of disease progression and not drug toxicity), and might be easier to measure. Nonetheless, bias because of patient heterogeneity and treatment decisions after disease progression might confound the results, but theoretically to a lesser extent than overall survival as a primary outcome measure.

Disease-free survival and event-free survival

In the setting of adjuvant therapy, in which cure is the aim, disease-free survival and event-free survival have been approved as surrogate endpoints in registered trials of oncology drugs.[9] The time needed to show an overall survival benefit is often impractical; therefore, disease-free survival and event-free survival are acceptable as surrogate endpoints and are clinically relevant. Both endpoints show the duration for which patients are disease free and are widely accepted as correlating with a better quality of life if not longer life.

Objective response rate

A surrogate endpoint often used for accelerated approval is the objective response rate.[8,9] Accelerated approval by regulatory bodies allows rapid access to drugs on the basis that the drug tested is better than current available therapy and that confirmatory trials to further investigate the potential actual benefit will be done.[34] In practice the duration of the response might be more clinically relevant than the extent of tumour reduction, particularly in the era of biological agents. Because many new drugs might be cytostatic rather than cytotoxic and result in stable disease, the objective response rate in patients might be low but a clear PFS and overall survival benefit can still be shown. Sorafenib in renal-cell cancer was associated with a less than 11% objective response rate, but a clear survival benefit.[35,36] Similarly in hepatocellular cancer, only 2% of patients treated with sorafenib had an objective response; however, a significant improvement was noted in median overall survival (10·7 months for sorafenib *vs* 7·9 months for placebo) and time to tumour progression (5·5 months *vs* 2·8 months).[37] In chemotherapy refractory non-small-cell lung cancer, erlotonib resulted in an objective response rate of only 8·9%, but a significant improvement in overall survival (6·7 months for erlotonib *vs* 4·7 months for placebo).[38]

The objective response rate includes the proportion of patients who have a partial and complete response and although any effect is directly attributable to the effect of the drug, the correlation between the response and clinical benefit can be unclear and subject to measurement error. Several biological agents have shown significant and rapid tumour response and objective response rates when matched to an appropriate biomarker, emphasising the importance of recognition of specific drug targets. For example, crizotinib gained FDA approval for use in patients with non-small-cell

lung cancer and the *ALK* fusion gene, based on results from two single-arm phase 2 studies[39,40] (255 patients) that showed objective response rates of 50%[39] and 61%.[40] By contrast, when gefitinib was studied in an unselected group of 1692 patients with advanced non-small-cell lung cancer, no significant benefit in overall survival was shown and drug development was reviewed.[41] *EGFR* mutation status (the biomarker of interest) was assessable in 215 patients in this study.[42] Only 26 (12%) of these patients had the biomarker of interest and on subsequent analysis a differential response rate was evident in patients with the biomarker (37·5% in those with an *EGFR* mutation *vs* 2·6% in those without).[42]

The importance of objective response rate as a surrogate endpoint in trials of cancer treatments is acknowledged by the FDA and used in breakthrough trials and for accelerated approval of drugs. Results of a study[43] of FDA drug approvals between 1990 and 2002, showed that 26 of 57 regular approvals were based on tumour response, with nine of these supported by improvement of tumour-related symptoms. Another study[44] showed similar findings. A high response rate in a patient population who have received previous treatments has been shown to be a reliable and robust criterion for drug approval.[45] Tsimberidou and colleagues[45] reported that one of 31 drugs approved on the basis of a single-arm study was withdrawn. Caution needs to be exercised as patients in a single-arm study might have fewer comorbidities, or have had better supportive care than in previous trials, which might inaccurately show the magnitude of any benefit.[45] Objective response rate in phase 2 trials has been associated with success in phase 3 trials.[46] A review of 89 trials of targeted therapies in six cancer types showed that objective response rate correlated with high non-progression rates and was predictive of eventual drug approval ($p=0·005$).[46] In metastatic breast cancer, achievement of an objective response in patients treated with chemotherapy predicted patient survival ($p<0·0001$),[47] with a similar relation noted in patients with advanced colorectal cancer.[48] Whether this association exists across cancer types and disease stages is as yet unknown.

Stable disease

Stable disease is not generally used as a criterion of response but is a guide for continuation of treatment.[49,50] A review of trials that contained a group of patients who did not receive any active treatment showed that disease stabilisation is reported in 25% (range 0–67%) of patients with metastatic solid tumours treated with placebo or best supportive care.[51] Many tumours that grow at their standard rate satisfy the definition of short-term stable disease. If disease stability is incorporated into trial design, it should only be done so if it is clinically meaningful or prolonged (eg, at least 6 months) and values should be interpreted with caution in slow-growing tumours.[51]

Quality of life

Health-related quality of life is defined as the effect of the illness on an individual's physical, psychological, social, and somatic functioning and general wellbeing.[52] Symptom control refers to alleviation of one or more symptoms and is complementary to health-related quality of life but should not be confused with it.[53] The FDA recognises symptomatic improvement as a direct clinical benefit to patients and an important consideration in drug approval.[8] Advances have been made with the classic measures of the European Organization for the Research and Treatment of Cancer (EORTC) Quality of life Questionnaire-Core 30 (QLQ-C30) and the Functional Assessment of Cancer Therapy, which now include more tumour, treatment and symptom-specific scores.[54–56] The assessment of health-related quality of life is a critical component of trials, at least as a secondary outcome, as a substantial proportion of systemic treatment is only given to achieve palliation of symptoms.

The importance of showing an improvement in quality of life is great, particularly if incremental gains in overall survival are small.[53] A study[57] of triplet therapy with 5-fluorouracil, oxaliplatin, and irinotecan compared with gemcitabine alone in pancreatic cancer showed the importance of health-related quality of life as an endpoint. Concerns about the survival benefit of triplet therapy balanced against the incidence of grade 3 or 4 drug toxicities were successfully addressed when it was shown that health-related quality of life was improved with triplet therapy.[58,59]

The interpretation of health-related quality of life remains a challenge as the assessment is by definition subjective and difficult to generalise between patient populations. Meaningful benefit to individual patients might be masked by the overall assessment of the group.[54] The ability of a patient to complete questionnaires might be biased depending on the timing of the assessments and when they take place in the trajectory of their illness (figure 2).[54] Changes in health-related quality of life might not be related to a patient's ongoing treatment, but rather to symptoms of disease progression or even the consequence of previous therapies. A study of patients with non-small-cell lung cancer showed greater health-related quality of life impairments with subsequent drug treatments (figure 2).[60] A patient given bad news is also more likely to interpret a questionnaire very differently from someone who has been told they are responding to therapy. Also, the drug toxicity and benefit deemed acceptable to a 30-year-old patient might be very different to that for an 80-year-old patient.

The assessment of health-related quality of life requires the same rigour as does the assessment of objective response rate and survival endpoints. A review[53] of 112 randomised controlled trials showed that the methods for assessing health-related quality of life differed

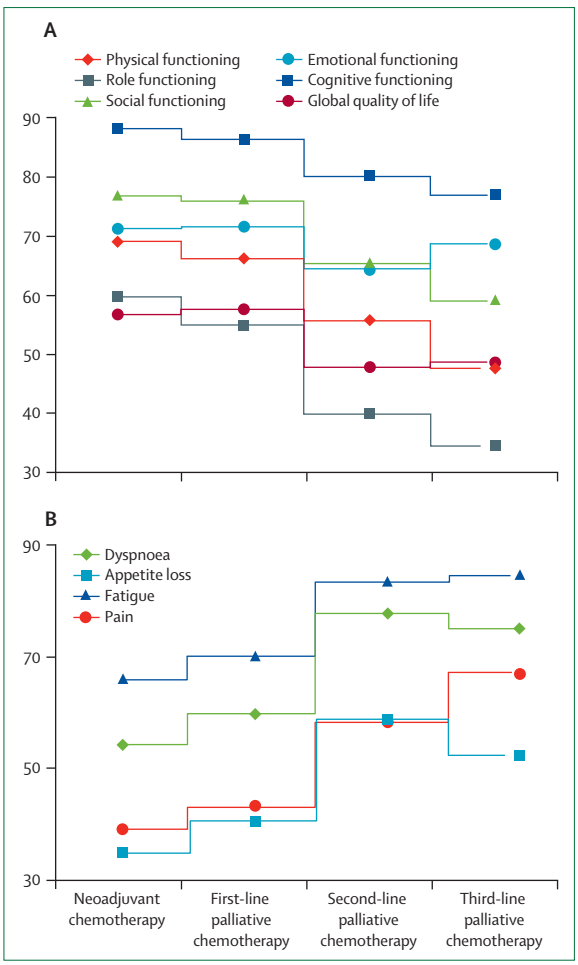

***Figure 2*: Functioning and global quality of life (A) and symptoms (B) in patients with non-small-cell lung cancer and different chemotherapy lines**
Patient functioning varies during disease course with generally decreased functioning with subsequent therapeutic lines.[60] Disease and treatment symptoms tend to increase with subsequent treatments. Reproduced from Wintner and colleagues,[60] by permission of Nature Publishing Group.

subtantially between studies. Although health-related quality of life is a property of the individual patient, only 25% of these studies reported individual patient scores.[53] Reporting of changes in group means or medians can mask individual therapeutic benefit. Only 21% of studies defined what constituted a response to palliative treatment, 13% reported on the duration of the response, and only 4% confirmed that a response to palliative treatment had occurred by reassessing the patient at a later date.[53] At present few published examples exist to show that a formal assessment of health-related quality of life or symptom control have been primarily used for drug funding decisions but these factors must be a priority component in trials of palliative treatment and are recognised as important by regulatory agencies.

There is a paucity of medical literature examining the relation between PFS and individual patient-reported outcomes, despite this being an essential aim of care. An exploratory analysis[19] of treatment with best supportive care with or without panitumumab in patients with metastatic colorectal cancer showed a significant association between health-related quality of life, disease symptoms, and PFS. The authors of this study postulated that clinical benefit was due to the halting of tumour growth with 10% of patients achieving a partial response and a further 27% with stable disease.[19] Similar effects have been shown with sorafenib in renal cancer,[34] and gemcitabine and capecitabine in biliary tract cancers.[61]

Technological implications

Accurate determination of clinical outcomes is dependent on the robustness of the methods used in their measurement. Clinical, biochemical, and radiological assessments of patients are used in day-to-day practice but traditionally in trials, responses in solid cancers, are based only on radiological measures in an attempt to ensure objectivity. Additionally, masked independent central and objective review of radiological findings is increasingly needed if the primary outcome is dependent on imaging criteria.[62] However, even this criterion can be associated with evaluation bias.[63]

Biochemical markers of disease progression have limited use in the determination of clinical outcomes and are rarely used independently of other assessments. The Gynecologic Cancer InterGroup has developed guidelines for monitoring of CA-125 concentrations (a tumour marker that might be elevated in some cancers) in the treatment of ovarian cancer but generally these are secondary outcomes and not used to define disease progression.[64,65] In prostate cancer, prostate specific antigen (PSA) is used as a marker of disease progression or response, often in conjunction with quality-of-life measures, because many patients have no quantifiable disease by radiological criteria.[66,67]

RECIST

In 1979 WHO established the first internationally recognised criteria for the radiological assessment of the response of solid malignancies to therapy that required 2D imaging and summation of tumour burden.[68,69] The objective Response Evaluation Criteria In Solid Tumours (RECIST) were established in 2000 (version 1.0)[68,70] and updated in 2009 (version 1.1).[20] By contrast with WHO criteria, RECIST used 1D radiological assessment and defined a partial response as a 30% reduction in the sum of maximum diameters of a specified number of measurable lesions.[20] These criteria were designed to provide an objective, uniform, and reliable method to show changes in tumour burden in trials between international institutions.[20,68,70]

Variability in the measurement of response remains problematic. A study[71] in non-small-cell lung cancer showed intra-observer misclassification of unchanged lesions as progressive disease in 9·5% of lesions and inter-observer misclassification of unchanged lesions as progressive disease in 29·8% of lesions. The challenge is

to show a correlation between the clinical status of the patient and the changes defined according to RECIST as progressive disease or partial response.[18] A change of 19% versus 21% in tumour size according to RECIST is unlikely to result in a change in symptoms, yet this is the defining criterion for tumour progression.[18] Furthermore, a patient with tumour regression of 29% might have a different outcome compared with a patient who has tumour growth of 19% but both would be labelled as stable disease according to RECIST. In practice it seems that patients with tumour regression are more likely to show symptomatic benefit than those without tumour regression. The use of tumour response as a continuous variable has been hypothesised to be more informative than specific cutoffs and to correlate better with patient survival.[72] However, studies have shown no improvement in the usefulness of assessing tumour response with this method to predict outcome when compared with RECIST measures.[73,74]

The timing of assessment must reflect the biological mechanism of treatment and the anticipated time to response as some tumours might seemingly progress after the start of treatment before a response is noted. A case in point is prostate cancer, in which PSA concentrations might initially increase after treatment and a bone scan might falsely show disease progression.[71] To account for this occurrence (known as flare) on bone scans, guidelines recommend delaying assessment of disease and PSA concentrations until after the first 12 weeks of therapy unless there are continual signs of progression.[75]

RECIST was established when the main therapy available was cytotoxic chemotherapy. Clinical experience has shown the limitations of these criteria to characterise the activity of new targeted treatments, biological treatments, and immunotherapy, largely because lesion volume on imaging might poorly show the number of viable tumour cells present.[76–78] For example, in gastrointestinal stromal tumours, assessment according to RECIST 1.0 can underestimate the response to imatinib.[72] A reduction in tumour density on CT of more than 15% (Hounsfield units) or a 10% decrease in tumour size (Choi criteria[79]) at 2 months was more sensitive and specific than was RECIST for assessment of response to treatment.[79]

Angiogenic drugs are often associated with cavitation and central necrosis of tumours making accurate assessment of clinical response and tumour progression a challenge.[80] Additionally, because many targeted therapies are cytostatic rather than cytotoxic, tumour shrinkage might not happen at the start of treatment.[81] Disease stabilisation, delayed response, or transient swelling followed by a clinical response are often frequent outcomes.[77] Similar results have been shown with immunotherapeutic agents such as ipilimumab in metastatic melanoma, in which tumour responses might occur after a period of tumour flare or pseudo-progression.[77,82–84] As a result, modified immune-response-related criteria[78] were created to optimise the assessment of tumour response or progression, but these criteria require validation and assessment as surrogate endpoints for overall survival.

Imaging advances

Differentiation of viable tumour tissue from treatment-induced fibrosis or necrosis on conventional CT is difficult.[68] A growing interest exists in ^{18}F-fluorodeoxy-glucose (^{18}F-FDG) PET and dynamic contrast enhanced MRI to distinguish active from inactive disease.

A PET scan after one to two cycles of chemotherapy has been shown to predict the response in lymphoma and is proposed to be a better marker of tumour response to some targeted drugs than is RECIST.[85–87] For example, in patients with non-small-cell lung cancer the response to gefitinib (but not chemotherapy) on PET was associated with a late anatomical response, PFS, and importantly, overall survival.[88] ^{18}F-FDG uptake is thought to correlate with the number of cancer cells and frequently decreases in response to treatment more quickly than a reduction in tumour size.[76,79,89,90] Integration of PET as a frontline method to assess the response of tumours to therapy in trials remains to be validated.[91] Not all diseases have high activity on ^{18}F-FDG PET and early changes in glucose uptake in tumours are not necessarily predictive of the response of all targeted treatments, as has been shown with mTOR inhibitors.[92]

The specificity and sensitivity of PET for detection of a change in tumour size will probably improve with new imaging reagents[87] and might improve disease staging and identify potential patients suitable for specific targeted treatment. In neuroendocrine tumours, somatostatin receptor imaging is used for staging and to identify patients who might benefit from treatment with radiolabelled analogues [^{90}Y] DOTATOC and [^{177}Lu] octreotate, and edotreotide.[93–95]

Circulating tumour cells and DNA

New indicators of disease progression continue to be identified, and might, in future, challenge traditional definitions of a clinical response. In some cancers, circulating tumour cells are postulated to provide real-time characterisation of disease status during therapy,[96] and have shown predictive and prognostic value among patients with metastatic breast, prostate, and colorectal cancers.[97–100] In metastatic breast cancer, circulating tumour cell quantification has been shown to be more reproducible and correlates better with patient survival than early radiological assessment (WHO criteria).[101] In prostate cancer circulating tumour cell counts predicted overall survival better than did decrements in PSA concentration ($p<0·02$).[98] Furthermore, circulating tumour cell counts can provide prognostic information, such as in colorectal cancer when used with imaging.[102] Circulating tumour cell counts have the advantage of allowing the assessment of response early, and potentially avoiding unnecessary

exposure to ineffective therapy. However, switching therapy in response to persistent increases of circulating tumour cell counts (at least in breast cancer) has not been shown to improve patient outcomes.[103] To date circulating tumour cell counts have not been validated in other cancers in which the pattern of spread is not mainly haematogenous.[104] In ovarian cancer a count of two circulating tumour cells per 7·5 mL of blood or higher was noted in only 14·4% of patients with relapsed disease.[104]

Circulating tumour DNA (ctDNA) might be more accessible and easier to process than circulating tumour cell counts.[105] In advanced cancer, ctDNA can constitute 1–10% of circulating DNA[106,107] and includes DNA shed by the primary tumour and from metastatic sites.[99] ctDNA concentration is thought to correlate with tumour burden because postoperative studies have shown that concentrations decrease after surgery and increase as new lesions arise.[108] Postoperative ctDNA concentration has shown better predictive value for recurrent disease than has carcinoembryonic antigen concentration in colorectal cancer.[108] In ovarian cancer, ctDNA concentration has been shown to parallel CA-125 concentration and disease activity.[107] Incorporation of circulating tumour cell counts and ctDNA concentration into large randomised controlled trials is restricted by the reproducibility of results between centres and the availability of the necessary technology.

Statistical considerations

Clinical trial outcomes are reflective of the appropriateness of the questions that are being asked. Preplanned statistical analyses provide an essential method to determine trial outcomes. Post-hoc analyses, although useful to generate hypotheses, are subject to bias. Statistics provide a formal framework to assess the strength of evidence,[109] but a significant result is still at risk of error. The probability of a research claim being true is dependent on study power and bias, the prior probability of it being true, and the statistical significance shown.[110] With more investigators asking similar questions, a high probability exists that an incorrect result might be achieved. With an α error of 0·05, one of 20 studies that ask the same question could be positive due to chance alone, independent of any true clinical benefit.[5] Results therefore require external validation and interpretation in parallel with other studies.[110]

Pivotal to trial design are examination of what is a plausible difference in response to treatment with respect to a measured clinical effect and what is the definition of a worthwhile clinical benefit. Some investigators suggest that large target benefits should be sought, thereby shifting current patient care practices.[111] A key advantage to a shift in standard therapy is that it would allow small studies and the potential for accelerated drug development. However, this advantage has to be balanced against a small sample size that could increase statistical uncertainty and the potential to overlook the small benefits that have to date shaped clinical practice.[111]

In oncology trials, p values remain the mainstay to assess whether the result was only due to chance. In most trials the p value is set at 0·05 as most clinicians are willing to accept a 5% probability of the result being a false positive. The question then arises as to whether there is a scenario in which clinicians would be willing to accept an increased risk of a false positive if more significant clinical benefit was achievable. At present, clinicians do accept an increased risk if the clinical benefits to the patient are substantial, particularly in situations in which the alternatives are limited (eg, the quick adoption of crizotinib in ALK-positive lung cancer based on phase 2 data alone).[39,40] Moreover, statistical simulations assessing two treatment superiority trials have shown that small sample sizes and high α values lead to increased long-term survival benefits compared with traditional trial designs.[4] These factors are particularly important as cancer becomes a collection of rare subentities with implications for trial recruitment and population size.[4]

Despite a trial result that is significant, gains in clinical terms can be small. The use of erlotinib in pancreatic cancer showed an 18% relative risk reduction (p=0·038), which translated into a less than 2-week improvement in median overall survival in patients.[112] The interpretation of statistics in the context of clinical value is important. The effect of drugs approved by the FDA on survival over the past 10 years has in general been small, with a median gain in survival of 2·16 months and PFS of 2·15 months.[113]

New horizons

Appropriate endpoints in trials depend on the clinical context, and require careful interpretation. The conundrum is how to intelligently use these endpoints to improve patient outcomes, define inter-tumour and inter-patient heterogeneity, and be clinically meaningful. Endpoints should ideally be of tangible benefit for patients, which in itself might be a challenge to show.

Quality-of-life scores need meticulous assessment and similar criteria and guidelines as used for clinical response to assess benefits, or lack of improvement, in quality of life.[53] The need to improve health-related quality of life scores and decrease drug toxicity in phase 2 and 3 clinical trials with palliative intent is fundamental and should be considered an important endpoint. The limiting aspect to this is the reproducibility and validity of present measures in practice.

Modest benefits in select patients in trials might decrease when the drug is used in the general community and in patients with comorbidities, perhaps even resulting in increased toxicity.[114,115] Randomised controlled trials are essential to detect or exclude a benefit to patients from a drug, but should be followed with observational or postmarketing studies to define the toxicity and effectiveness in the general community.[114,115]

The objective assessment of meaningful benefit for a patient is paramount in designing trials and measured outcomes need to be tailored to the patient population,

Search strategy and selection criteria

References were identified through searches of PubMed with the search terms "endpoints", "outcomes", "oncology", and "trial design". Only papers published in English were reviewed without any date restrictions. Articles were also identified through searches of the authors' own files and cited articles. The final reference list was generated on the basis of originality and relevance to the broad scope of this topic.

disease, and therapeutic modality. Trials and endpoints have to evolve to reflect the increased understanding of disease biology and predictors of response and benefit—how clinicians approach and bridge this divide is the next challenge and is explored in a companion Review.[116]

Contributors

MKW did the initial draft and medical literature search. The final version was reviewed and edited by MKW, KK, and AMO.

Declaration of interests

We declare no competing interests.

References

1 Booth CM, Tannock I. Reflections on medical oncology: 25 years of clinical trials—where have we come and where are we going? *J Clin Oncol* 2008; **26:** 6–8.

2 Zhuang SH, Xiu L, Elsayed YA. Overall survival: a gold standard in search of a surrogate: the value of progression-free survival and time to progression as end points of drug efficacy. *Cancer J* 2009; **15:** 395–400.

3 Stewart DJ, Whitney SN, Kurzrock R. Equipoise lost: ethics, costs, and the regulation of cancer clinical research. *J Clin Oncol* 2010; **28:** 2925–35.

4 Deley MC, Ballman KV, Marandet J, Sargent D. Taking the long view: how to design a series of phase III trials to maximize cumulative therapeutic benefit. *ClinTrials* 2012; **9:** 283–92.

5 Stewart DJ, Kurzrock R. Fool's gold, lost treasures, and the randomized clinical trial. *BMC Cancer* 2013; **13:** 193.

6 McNamara MG, Metran-Nascente C, Knox JJ. State-of-the-art in the management of locally advanced and metastatic gallbladder cancer. *Curr Opin Oncol* 2013; **25:** 425–31.

7 Rosario R, Wilson M, Cheng WT, et al. Adult granulosa cell tumours (GCT): clinicopathological outcomes including FOXL2 mutational status and expression. *Gynecol Oncol* 2013; **131:** 325–29.

8 Pazdur R. Endpoints for assessing drug activity in clinical trials. *Oncologist* 2008; **13** (suppl 2)**:** 19–21.

9 McKee AE, Farrell AT, Pazdur R, Woodcock J. The role of the US Food and Drug Administration review process: clinical trial endpoints in oncology. *Oncologist* 2010; **15** (suppl 1)**:** 13–18.

10 CHMP Scientific Advisory Group (SAG) for Oncology. Answers from the CHMP Scientific Advisory Group (SAG) for Oncology for Revision of the anticancer guideline. http://www.ema.europa.eu/docs/en_GB/document_library/Other/2013/01/WC500137129.pdf. (accessed Dec 17, 2013).

11 Driscoll JJ, Rixe O. Overall survival: still the gold standard: why overall survival remains the definitive end point in cancer clinical trials. *Cancer J* 2009; **15:** 401–05.

12 Baker SG, Kramer BS. Surrogate endpoint analysis: an exercise in extrapolation. *J Natl Cancer Inst* 2013; **105:** 316–20.

13 Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin Pharmacol Ther* 2001; **69:** 89–95.

14 Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria. *Stat Med* 1989; **8:** 431–40.

15 Baker SG, Sargent DJ, Buyse M, Burzykowski T. Predicting treatment effect from surrogate endpoints and historical trials: an extrapolation involving probabilities of a binary outcome or survival to a specific time. *Biometrics* 2012; **68:** 248–57.

16 Seruga B, Sterling L, Wang L, Tannock IF. Reporting of serious adverse drug reactions of targeted anticancer agents in pivotal phase III clinical trials. *J Clin Oncol* 2011; **29:** 174–85.

17 Matulonis UA1, Oza AM, Ho TW, Ledermann JA. Intermediate clinical endpoints: a bridge between progression-free survival and overall survival in ovarian cancer trials. *Cancer* 2014; published online Oct 21. DOI: 10.1002/cncr.29082.

18 Booth CM, Eisenhauer EA. Progression-free survival: meaningful or simply measurable? *J Clin Oncol* 2012; **30:** 1030–33.

19 Siena S, Peeters M, Van Cutsem E, et al. Association of progression-free survival with patient-reported outcomes and survival: results from a randomised phase 3 trial of panitumumab. *Br J Cancer* 2007; **97:** 1469–74.

20 Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer* 2009; **45:** 228–47.

21 Broglio KR, Berry DA. Detecting an overall survival benefit that is derived from progression-free survival. *J Natl Cancer Inst* 2009; **101:** 1642–49.

22 Buyse M, Burzykowski T, Carroll K, et al. Progression-free survival is a surrogate for survival in advanced colorectal cancer. *J Clin Oncol* 2007; **25:** 5218–24.

23 Tang PA, Bentzen SM, Chen EX, Siu LL. Surrogate end points for median overall survival in metastatic colorectal cancer: literature-based analysis from 39 randomized controlled trials of first-line chemotherapy. *J Clin Oncol* 2007; **25:** 4562–68.

24 Foster NR, Qi Y, Shi Q, et al. Tumor response and progression-free survival as potential surrogate endpoints for overall survival in extensive stage small-cell lung cancer: findings on the basis of North Central Cancer Treatment Group trials. *Cancer* 2011; **117:** 1262–71.

25 Sherrill B, Kaye JA, Sandin R, Cappelleri JC, Chen C. Review of meta-analyses evaluating surrogate endpoints for overall survival in oncology. *Onco Targets Ther* 2012; **5:** 287–96.

26 Slamon DJ, Leyland-Jones B, Shak S, et al. Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *N Engl J Med* 2001; **344:** 783–92.

27 Baselga J, Cortes J, Kim SB, et al. Pertuzumab plus trastuzumab plus docetaxel for metastatic breast cancer. *N Engl J Med* 2012; **366:** 109–19.

28 Temel JS, Greer JA, Muzikansky A, et al. Early palliative care for patients with metastatic non-small-cell lung cancer. *N Engl J Med* 2010; **363:** 733–42.

29 Sridhara R, Mandrekar SJ, Dodd LE. Missing data and measurement variability in assessing progression-free survival endpoint in randomized clinical trials. *Clin Cancer Res* 2013; **19:** 2613–20.

30 Baselga J, Campone M, Piccart M, et al. Everolimus in postmenopausal hormone-receptor-positive advanced breast cancer. *N Engl J Med* 2012; **366:** 520–29.

31 Ledermann J, Harter P, Gourley C, et al. Olaparib maintenance therapy in patients with platinum-sensitive relapsed serous ovarian cancer: a preplanned retrospective analysis of outcomes by *BRCA* status in a randomised phase 2 trial. *Lancet Oncol* 2014; **15:** 852–61.

32 Ledermann J, Harter P, Gourley C, et al. Olaparib maintenance therapy in platinum-sensitive relapsed ovarian cancer. *N Engl J Med* 2012; **366:** 1382–92.

33 Gelmon KA, Tischkowitz M, Mackay H, et al. Olaparib in patients with recurrent high-grade serous or poorly differentiated ovarian carcinoma or triple-negative breast cancer: a phase 2, multicentre, open-label, non-randomised study. *Lancet Oncol* 2011; **12:** 852–61.

34 Johnson JR, Ning YM, Farrell A, Justice R, Keegan P, Pazdur R. Accelerated approval of oncology products: the food and drug administration experience. *J Natl Cancer Inst* 2011; **103:** 636–44.

35 Escudier B, Eisen T, Stadler WM, et al. Sorafenib for treatment of renal cell carcinoma: final efficacy and safety results of the phase III treatment approaches in renal cancer global evaluation trial. *J Clin Oncol* 2009; **27:** 3312–18.

36 Escudier B, Eisen T, Stadler WM, et al. Sorafenib in advanced clear-cell renal-cell carcinoma. *N Engl J Med* 2007; **356:** 125–34.

37 Llovet JM, Ricci S, Mazzaferro V, et al. Sorafenib in advanced hepatocellular carcinoma. *N Engl J Med* 2008; **359:** 378–90.

38 Shepherd FA, Rodrigues Pereira J, Ciuleanu T, et al. Erlotinib in previously treated non-small-cell lung cancer. *N Engl J Med* 2005; **353:** 123–32.

39 Kim D, W, Ahn MJ, Shi Y, et al. Results of a global phase II study with crizotinib in advanced ALK-positive non-small cell lung cancer (NSCLC). *Proc Am Soc Clin Oncol* 2012; **30** (suppl 1)**:** abstr 7533.

40 Crinò LK, Riely GJ, Janne PA, et al. Initial phase II results with crizotinib in advanced ALK-positive non-small cell lung cancer (NSCLC): PROFILE 1005. *Proc Am Soc Clin Oncol* 2011; **29** (suppl)**:** abstr 7514.

41 Thatcher N, Chang A, Parikh P, et al. Gefitinib plus best supportive care in previously treated patients with refractory advanced non-small-cell lung cancer: results from a randomised, placebo-controlled, multicentre study (Iressa Survival Evaluation in Lung Cancer). *Lancet* 2005; **366:** 1527–37.

42 Hirsch FR, Varella-Garcia M, Bunn PA, Jr, et al. Molecular predictors of outcome with gefitinib in a phase III placebo-controlled study in advanced non-small-cell lung cancer. *J Clin Oncol* 2006; **24:** 5034–42.

43 Johnson JR, Williams G, Pazdur R. End points and United States Food and Drug Administration approval of oncology drugs. *J Clin Oncol* 2003; **21:** 1404–11.

44 Martell RE, Sermer D, Getz K, Kaitin KI. Oncology drug development and approval of systemic anticancer therapy by the U.S. Food and Drug Administration. *Oncologist* 2013; **18:** 104–11.

45 Tsimberidou AM, Braiteh F, Stewart DJ, Kurzrock R. Ultimate fate of oncology drugs approved by the us food and drug administration without a randomized trial. *J Clin Oncol* 2009; **27:** 6243–50.

46 El-Maraghi RH, Eisenhauer EA. Review of phase II trial designs used in studies of molecular targeted agents: outcomes and predictors of success in phase III. *J Clin Oncol* 2008; **26:** 1346–54.

47 Bruzzi P, Del Mastro L, Sormani MP, et al. Objective response to chemotherapy as a potential surrogate end point of survival in metastatic breast cancer patients *J Clin Oncol* 2005; **23:** 5117–25.

48 Buyse M, Thirion P, Carlson RW, Burzykowski T, Molenberghs G, Piedbois P. Relation between tumour response to first-line chemotherapy and survival in advanced colorectal cancer: a meta-analysis. Meta-Analysis Group in Cancer. *Lancet* 2000; **356:** 373–78.

49 Tolcher AW. Stable disease is a valid end point in clinical trials. *Cancer J* 2009; **15:** 374–48.

50 Vidaurre T, Wilkerson J, Simon R, Bates SE, Fojo T. Stable disease is not preferentially observed with targeted therapies and as currently defined has limited value in drug development. *Cancer J* 2009; **15:** 366–73.

51 Le Tourneau C, Paoletti X, Coquan E, Sablin MP, Zoubir M, Tannock IF. Critical evaluation of disease stabilization as a measure of activity of systemic therapy: lessons from trials with arms in which patients do not receive active treatment. *J Clin Oncol* 2014; **32:** 260–63.

52 Revicki DA, Osoba D, Fairclough D, et al. Recommendations on health-related quality of life research to support labeling and promotional claims in the United States. *Qual Life Res* 2000; **9:** 887–900.

53 Joly F, Vardy J, Pintilie M, Tannock IF. Quality of life and/or symptom control in randomized clinical trials for patients with advanced cancer. *Ann Oncol* 2007; **18:** 1935–42.

54 Trask PC, Hsu MA, McQuellon R. Other paradigms: health-related quality of life as a measure in cancer treatment: its importance and relevance. *Cancer J* 2009; **15:** 435–40.

55 Aaronson NK, Ahmedzai S, Bergman B, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. *J Natl Cancer Inst* 1993; **85:** 365–76.

56 Sullivan R. Clinical trial design in oncology: protocol design. *Lancet Oncol* 2004; **5:** 759.

57 Conroy T, Desseigne F, Ychou M, et al. FOLFIRINOX versus gemcitabine for metastatic pancreatic cancer. *N Engl J Med* 2011; **364:** 1817–25.

58 Ko AH. FOLFIRINOX: a small step or a great leap forward? *J Clin Oncol* 2011; **29:** 3727–29.

59 Gourgou-Bourgade S, Bascoul-Mollevi C, Desseigne F, et al. Impact of FOLFIRINOX compared with gemcitabine on quality of life in patients with metastatic pancreatic cancer: results from the PRODIGE 4/ACCORD 11 randomized trial. *J Clin Oncol* 2013; **31:** 23–29.

60 Wintner LM, Giesinger JM, Zabernigg A, et al. Quality of life during chemotherapy in lung cancer patients: results across different treatment lines. *Br J Cancer* 2013; **109:** 2301–08.

61 Koeberle D, Saletti P, Borner M, et al. Patient-reported outcomes of patients with advanced biliary tract cancers receiving gemcitabine plus capecitabine: a multicenter, phase II trial of the Swiss Group for Clinical Cancer Research. *J Clin Oncol* 2008; **26:** 3702–08.

62 Ford R, Schwartz L, Dancey J, et al. Lessons learned from independent central review. *Eur J Cancer* 2009; **45:** 268–64.

63 Amit O, Bushnell W, Dodd L, Roach N, Sargent D. Blinded independent central review of the progression-free survival endpoint. *Oncologist* 2010; **15:** 492–95.

64 Rustin GJ, Quinn M, Thigpen T, et al. Re: new guidelines to evaluate the response to treatment in solid tumors (ovarian cancer). *J Natl Cancer Inst* 2004; **96:** 487–88.

65 Rustin GJ, Vergote I, Eisenhauer E, et al. Definitions for response and progression in ovarian cancer clinical trials incorporating RECIST 1.1 and CA 125 agreed by the Gynecological Cancer Intergroup (GCIG). *Int J Gynecol Cancer* 2011; **21:** 419–23.

66 de Bono JS, Oudard S, Ozguroglu M, et al. Prednisone plus cabazitaxel or mitoxantrone for metastatic castration-resistant prostate cancer progressing after docetaxel treatment: a randomised open-label trial. *Lancet* 2010; **376:** 1147–54.

67 Tannock IF, de Wit R, Berry WR, et al. Docetaxel plus prednisone or mitoxantrone plus prednisone for advanced prostate cancer. *N Engl J Med* 2004; **351:** 1502–12.

68 Bradbury P, Seymour L. Tumor shrinkage and objective response rates: gold standard for oncology efficacy screening trials, or an outdated end point? *Cancer J* 2009; **15:** 354–60.

69 Miller AB, Hoogstraten B, Staquet M, Winkler A. Reporting results of cancer treatment. *Cancer* 1981; **47:** 207–14.

70 Therasse P, Arbuck SG, Eisenhauer EA, et al. New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada. *J Natl Cancer Inst* 2000; **92:** 205–16.

71 Erasmus JJ, Gladish GW, Broemeling L, et al. Interobserver and intraobserver variability in measurement of non-small-cell carcinoma lung lesions: implications for assessment of tumor response. *J Clin Oncol* 2003; **21:** 2574–82.

72 Jain RK, Lee JJ, Ng C, et al. Change in tumor size by RECIST correlates linearly with overall survival in phase I oncology studies. *J Clin Oncol* 2012; **30:** 2684–90.

73 An MW, Mandrekar SJ, Branda ME, et al. Comparison of continuous versus categorical tumor measurement-based metrics to predict overall survival in cancer treatment trials. *Clin Cancer Res* 2011; **17:** 6592–99.

74 Mandrekar SJ, An MW, Meyers J, Grothey A, Bogaerts J, Sargent DJ. Evaluation of alternate categorical tumor metrics and cut points for response categorization using the RECIST 1.1 data warehouse. *J Clin Oncol* 2014; **32:** 841–50.

75 Scher HI, Morris MJ, Basch E, Heller G. End points and outcomes in castration-resistant prostate cancer: from clinical trials to clinical practice. *J Clin Oncol* 2011; **29:** 3695–704.

76 Choi H, Charnsangavej C, de Castro Faria S, et al. CT evaluation of the response of gastrointestinal stromal tumors after imatinib mesylate treatment: a quantitative analysis correlated with FDG PET findings. *AJR Am J Roentgenol* 2004; **183:** 1619–28.

77 Hales RK, Banchereau J, Ribas A, et al. Assessing oncologic benefit in clinical trials of immunotherapy agents. *Ann Oncol* 2010; **21:** 1944–51.

78 Wolchok JD, Hoos A, O'Day S, et al. Guidelines for the evaluation of immune therapy activity in solid tumors: immune-related response criteria. *Clin Cancer Res* 2009; **15:** 7412–20.

79 Choi H, Charnsangavej C, Faria SC, et al. Correlation of computed tomography and positron emission tomography in patients with metastatic gastrointestinal stromal tumor treated at a single institution with imatinib mesylate: proposal of new computed tomography response criteria. *J Clin Oncol* 2007; **25:** 1753–59.

80 Crabb SJ, Patsios D, Sauerbrei E, et al. Tumor cavitation: impact on objective response evaluation in trials of angiogenesis inhibitors in non-small-cell lung cancer. *J Clin Oncol* 2009; **27:** 404–10.

81 Ratain MJ, Eisen T, Stadler WM, et al. Phase II placebo-controlled randomized discontinuation trial of sorafenib in patients with metastatic renal cell carcinoma. *J Clin Oncol* 2006; **24:** 2505–12.

82 Nishino M, Giobbie-Hurder A, Gargano M, Suda M, Ramaiya NH, Hodi FS. Developing a common language for tumor response to immunotherapy: immune-related response criteria using unidimensional measurements. *Clin Cancer Res* 2013; **19:** 3936–43.

83 Saenger YM, Wolchok JD. The heterogeneity of the kinetics of response to ipilimumab in metastatic melanoma: patient cases. *Cancer Immun* 2008; **8:** 1.

84 O'Day SJ, Hamid O, Urba WJ. Targeting cytotoxic T-lymphocyte antigen-4 (CTLA-4): a novel strategy for the treatment of melanoma and other malignancies. *Cancer* 2007; **110:** 2614–27.

85 Kasamon YL, Wahl RL. FDG PET and risk-adapted therapy in Hodgkin's and non-Hodgkin's lymphoma. *Curr Opin Oncol* 2008; **20:** 206–19.

86 Wahl RL, Jacene H, Kasamon Y, Lodge MA. From RECIST to PERCIST: evolving considerations for PET response criteria in solid tumors. *J Nucl Med* 2009; **50** (suppl 1)**:** 122–50.

87 Weber WA. Positron emission tomography as an imaging biomarker. *J Clin Oncol* 2006; **24:** 3282–92.

88 Kanazu M, Maruyama K, Ando M, et al. Early pharmacodynamic assessment using f-fluorodeoxyglucose positron-emission tomography on molecular targeted therapy and cytotoxic chemotherapy for clinical outcome prediction. *Clin Lung Cancer* 2014; **15:** 182–87.

89 Bos R, van Der Hoeven JJ, van Der Wall E, et al. Biologic correlates of ^{18}fluorodeoxyglucose uptake in human breast cancer measured by positron emission tomography. *J Clin Oncol* 2002; **20:** 379–87.

90 Schoder H, Fury M, Lee N, Kraus D. PET monitoring of therapy response in head and neck squamous cell carcinoma. *J Nucl Med* 2009; **50** (suppl 1)**:** 74–88.

91 Sargent DJ, Rubinstein L, Schwartz L, et al. Validation of novel imaging methodologies for use as cancer clinical trial end-points. *Eur J Cancer* 2009; **45:** 290–99.

92 Ma WW, Jacene H, Song D, et al. ^{18}F fluorodeoxyglucose positron emission tomography correlates with Akt pathway activity but is not predictive of clinical outcome during mTOR inhibitor therapy. *J Clin Oncol* 2009; **27:** 2697–704.

93 Waldherr C, Pless M, Maecke HR, Haldemann A, Mueller-Brand J. The clinical value of [^{90}Y-DOTA]-D-Phe^{1}-Tyr^{3}-octreotide (^{90}Y-DOTATOC) in the treatment of neuroendocrine tumours: a clinical phase II study. *Ann Oncol* 2001; **12:** 941–45.

94 Kwekkeboom DJ, Teunissen JJ, Bakker WH, al. Radiolabeled somatostatin analog [^{177}Lu-DOTA0,Tyr^{3}]octreotate in patients with endocrine gastroenteropancreatic tumors. *J Clin Oncol* 2005; **23:** 2754–62.

95 Kwekkeboom DJ, Kam BL, van Essen M, et al. Somatostatin-receptor-based imaging and therapy of gastroenteropancreatic neuroendocrine tumors. *Endocr Relat Cancer* 2010; **17:** R53–73.

96 Attard G, Swennenhuis JF, Olmos D, et al. Characterization of *ERG*, *AR* and *PTEN* gene status in circulating tumor cells from patients with castration-resistant prostate cancer. *Cancer Res* 2009; **69:** 2912–18.

97 Cristofanilli M, Budd GT, Ellis MJ, et al. Circulating tumor cells, disease progression, and survival in metastatic breast cancer. *N Engl J Med* 2004; **351:** 781–91.

98 de Bono JS, Scher HI, Montgomery RB, et al. Circulating tumor cells predict survival benefit from treatment in metastatic castration-resistant prostate cancer. *Clin Cancer Res* 2008; **14:** 6302–09.

99 Alix-Panabieres C, Schwarzenbach H, Pantel K. Circulating tumor cells and circulating tumor DNA. *Annu Rev Med* 2012; **63:** 199–215.

100 Scher HI, Jia X, de Bono JS, et al. Circulating tumour cells as prognostic markers in progressive, castration-resistant prostate cancer: a reanalysis of IMMC38 trial data. *Lancet Oncol* 2009; **10:** 233–39.

101 Budd GT, Cristofanilli M, Ellis MJ, et al. Circulating tumor cells versus imaging—predicting overall survival in metastatic breast cancer. *Clin Cancer Res* 2006; **12:** 6403–09.

102 Cohen SJ, Punt CJ, Iannotti N, et al. Relationship of circulating tumor cells to tumor response, progression-free survival, and overall survival in patients with metastatic colorectal cancer. *J Clin Oncol* 2008; **26:** 3213–21.

103 Smerage JB, Barlow WE, Hortobagyi GN, et al. Circulating tumor cells and response to chemotherapy in metastatic breast cancer: SWOG S0500. *J Clin Oncol* 2014; published online June 2. DOI:10.1200/JCO.2014.56.2561.

104 Poveda A, Kaye SB, McCormack R, et al. Circulating tumor cells predict progression free survival and overall survival in patients with relapsed/recurrent advanced ovarian cancer. *Gynecol Oncol* 2011; **122:** 567–72.

105 Murtaza M, Dawson SJ, Tsui DW, et al. Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA. *Nature* 2013; **497:** 108–12.

106 Diehl F, Li M, Dressman D, et al. Detection and quantification of mutations in the plasma of patients with colorectal tumors. *Proc Natl Acad Sci USA* 2005; **102:** 16368–73.

107 Forshew T, Murtaza M, Parkinson C, et al. Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA. *Sci Trans Med* 2012; **4:** 136ra68.

108 Diehl F, Schmidt K, Choti MA, et al. Circulating mutant DNA to assess tumor dynamics. *Nat Med* 2008; **14:** 985–90.

109 Lee JJ. Demystify statistical significance: time to move on from the p value to bayesian analysis. *J Natl Cancer Inst* 2011; **103:** 2–3.

110 Ioannidis JP. Why most published research findings are false. *PLoS Med* 2005; **2:** e124.

111 Sobrero A, Bruzzi P. Incremental advance or seismic shift? The need to raise the bar of efficacy for drug approval. *J Clin Oncol* 2009; **27:** 5868–73.

112 Moore MJ, Goldstein D, Hamm J, et al. Erlotinib plus gemcitabine compared with gemcitabine alone in patients with advanced pancreatic cancer: a phase III trial of the National Cancer Institute of Canada Clinical Trials Group. *J Clin Oncol* 2007; **25:** 1960–66.

113 Fojo AT, Noonan A. Why RECIST works and why it should stay—counterpoint. *Cancer Res* 2012; **72:** 5151–57.

114 Booth CM, Tannock IF. Randomised controlled trials and population-based observational research: partners in the evolution of medical evidence. *Br J Cancer* 2014; **110:** 551–55.

115 Booth CM, Tannock IF. Evaluation of treatment benefit: randomized controlled trials and population-based observational research. *J Clin Oncol* 2013; **31:** 3298–99.

116 Wilson MK, Collyar D, Chingos DT, et al. Outcomes and endpoints in cancer trials: bridging the divide. *Lancet Oncol* 2014; **16:** e43–52.