Clinical trial end points relevant to patients and society for rare cancers

In solid tumors, end points such as progression-free survival are increasingly utilized as primary end points, as the use of overall survival can often be confounded by the growing use of multiple lines of therapy. In rare cancers, the choice of end points is further complicated by small and heterogeneous patient populations. In the absence of confirmed overall survival benefit, it remains unclear as to whether extending progression-free survival provides a discernible clinical benefit. Inclusion of robust patient-reported outcomes may provide valuable supporting evidence when making decisions regarding the clinical value of new costly agents. We discuss recent trials in pancreatic neuroendocrine tumors to exemplify some of the challenges faced in the trial design for rare cancers.

Keywords: end points • health related quality of life • objective response rate • overall survival • pancreatic neuroendocrine tumors • progression-free survival • rare cancers

There is no internationally agreed definition of a rare cancer. However, based on the RARECARE working group definition, the incidence of rare cancers is <6 per 100,000 per year [1]. Taken together, rare cancers account for 22% of all cancer diagnoses, including all cancers in children. The combined incidence of all rare cancers is actually higher than any of the individual common cancers and therefore rare cancers should be regarded as a significant public health problem.

The study of rare cancers is faced with unique challenges. Patients are often misdiagnosed or diagnosed late [2]. There is a scarcity of evidence to help guide treatment decisions and often lack of expertise among treating clinicians. Moreover, industry often prioritizes cancers with a larger potential market. Academic research is limited by the small patient populations which compromises any attempts of undertaking single institution trials that would carry adequate statistical power. In recognition of this, international collaborations such as the International Rare Cancers Initiative have recently been developed to facilitate the development of international clinical trials [2].

In this review, we will discuss the strengths and limitations of end points used in trials of common cancers and then use pancreatic neuroendocrine tumors (pNETs) as an example of a rare cancer to illustrate some of the more specific issues encountered when designing trials for such patient populations. We will discuss recent trials that have been undertaken in this tumor group, focusing on the study methodology and the end points that were employed, in order to exemplify the challenges faced when designing trials for rare cancers.

A discussion of the strengths & limitations of end points used in oncology trials

Optimal clinical trial end points have long been debated in oncology. While it is well established that overall survival (OS) is the gold standard due to its unambiguous and objective nature, it is increasingly recognized that it does have its own limitations. OS is confounded by the growing use of multiple lines of therapy, and consequently, the use of progression-free survival (PFS) as an end point is increasingly common. PFS is generally easier to measure, less confounded by salvage treatments, and may be considered a more meaningful endpoint to patients. However, PFS is not without its own limitations, particularly in small and heterogeneous patient populations where the benefits of treatment may not be apparent.

There are additional end points that may provide valuable supporting evidence when making decisions regarding the clinical value of new costly agents. These include patient-reported outcomes, which may provide valuable information regarding the quality of life and symptoms experienced by patients. However, these outcomes must be rigorously validated and standardized to ensure comparability across studies.

The choice of end points is crucial in the design of clinical trials, and it is important to consider the potential limitations of each end point when making decisions regarding the design of future trials. The use of multiple end points may provide a more comprehensive understanding of the clinical benefit of new agents, but it is important to ensure that these end points are appropriately chosen and standardized to ensure comparability across studies.

Acknowledgements

This work was supported by the National Institute for Health Research (NIHR) Biomedical Research Centre at the Royal Marsden Hospital and Institute of Cancer Research. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health.

Conflict of interest

The authors declare that there is no conflict of interest.

References

facts-at-a-glance.[Accessed 15 December 2015].
lines of therapy and is often associated with prolonged follow-up with higher attrition rates and larger sample sizes that are required to reach statistical significance.

In view of the aforementioned issues, along with the high failure rates in the latter stages of drug development (with only one in 20 cancer drugs achieving US FDA approval) in a field of medicine where new treatment paradigms are needed, recent clinical trials have utilized surrogate end points as a substitute for OS (3). This is demonstrated by a recent publication which found that between 2002 and 2012, up to two thirds of FDA oncology regular approvals were based on end points other than OS (4).

The most commonly used surrogate end points in the metastatic setting are progression-free survival (PFS), time to progression (TTP) and objective response rate (ORR). PFS is defined as the time from randomization until objective tumor progression or death and TTP is defined as time from randomization until objective tumor progression, not including death (see Table 1). Unlike OS, all three end points have the advantage that they are not affected by subsequent lines of treatment or crossover to the experimental arm. However, in unblinded trials, the use of any of these primary end points can introduce bias, as identifying radiological progression has a subjective element. Therefore, investigators may be slower to declare progression in the investigative medicinal product arm and similarly, patients in the control group may be quicker to report symptoms associated with progression. This source of bias can be minimized by the inclusion of blinded independent central review (BICR), but this approach is associated with increased potential for measurement variability, cost and complexity of trial design (5). Furthermore, if BICR is not done in real time, there is a risk for patients and clinicians to withdraw from a trial early with subsequent BICR declaring lack of radiological progression some time after trial withdrawal. This then further complicates interpretation of clinical trial results as was the case in the RADIANT 2 study (6). A meta-analysis based on 27 randomized Phase III studies found a strong correlation between local evaluation and BICR and advised that in cases where a trial is blinded or there is a large observed effect on PFS, BICR may not be necessary. They suggested that where BICR is warranted, such as in smaller trials, a sample-based BICR may increase trial credibility without significantly affecting trial complexity and cost (7). If there is discordance between the sample-based BICR and local evaluation, BICR may then be justified for the whole trial population.

Apart from the subjectivity associated with assessing radiological progression, the timing of assessments also influences PFS. Trial protocols may pre-specify when patients should have radiological response assessments performed. If evidence of progressive disease is identified, there is no way of determining the exact progression date. Progression could have occurred at any time point from the previous scan leading up to the latest scan. This phenomenon could lead to overestimation of PFS and has been highlighted in a previous study, where the authors propose that increasing the use of specific statistical methods to analyze interval-censored data may help minimize this source of bias (8). They also suggest that there should be some consistency in the timing of response assessments among studies within a particular tumor site. While they recognize that cross trial comparisons are discouraged, consistency in the timing of response assessments would facilitate the interpretation of PFS results from a number of studies within the same field.

The Response Evaluation Criteria in Solid tumors were developed to help standardize and unify response assessment between trials and to date remain the most commonly utilized criteria in clinical trials (9). However, since the establishment of these criteria, clinicians and investigators alike have come to recognize some of their limitations (10). The response thresholds of a 30% decrease for response and a 20% increase for progression were set arbitrarily without any evidence to support that these particular thresholds affect patient outcomes. While cytotoxic agents can lead to significant changes in the size of target lesions, newer cytostatic targeted agents may not demonstrate such dramatic changes but may still be beneficial to patients by providing some disease control in conjunction with improvement in symptoms. Moreover, the use of newer immunotherapy drugs can lead to initial tumor flare prior to inducing a response and therefore assessment by Response Evaluation Criteria in Solid tumors may incorrectly term this disease progression. Ideally, response assessment criteria need to be treatment and disease specific. In recognition of this, specific criteria have been developed for mesothelioma, hepatocellular carcinoma and immunotherapy to name but a few (11–13). Rare cancers have not been entirely neglected in the evolving response assessment paradigm. Imatinib, a selective inhibitor of the kinase activity of KIT and platelet-derived growth factor receptor, is an effective treatment for gastrointestinal stromal tumor (GIST), a rare cancer. Imatinib is known to induce tumor necrosis and therefore paradoxically can increase the size of target lesions, appearing as progression on imaging. The Choi criteria were devised to account for this (14). The use of these criteria for other tumor types with other targeted therapies has, however, yielded conflicting results (15–47). The development and validation of new treatment specific or cancer specific response
assessment criteria is an interesting and rapidly evolving field and highlights the importance of considering the limitations of any response assessment method, particularly when the primary end point of a trial is tumor centered.

The use of surrogate end points such as PFS has frequently been criticized due to a lack of consistency of correlation with OS across all tumor types (see Table 2) and the variation in the definitions used across trials [18]. Differing definitions for end points are highlighted by the initial definition of clinical benefit as a composite assessment of pain, performance status and weight in a landmark study of gemcitabine in pancreatic cancer [19]. With the increasing use of targeted therapy, there has been a shift away from this definition to a more tumor-centered description incorporating tumor response and stability. Extending or improving quality of life might usually be what one might consider the term clinical benefit to reflect. However, the tumor-centered definition of clinical benefit that we have commonly come to accept does not consistently correlate with these measures. It has therefore been suggested that disease control rate would be a more accurate description of the tumor-centered definition of clinical benefit [20].

Tumor-centered end points in isolation such as PFS do not provide a direct clinical benefit for patients. In particular, PFS seems to be a poor surrogate for OS in patients with a long survival post-progression (SPP) compared with those with a shorter SPP [40,41]. It has been suggested that in tumors with long SPP, PFS is only clinically meaningful when it is also associated

<table>
<thead>
<tr>
<th>End point</th>
<th>Definition</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>OS</strong></td>
<td>Time from randomization to death from any cause</td>
<td>Unequivocal, objective, easy to interpret</td>
<td>Requires larger sample size, longer follow-up which can be associated with higher attrition rate, more costly, confounded by subsequent lines of treatment</td>
</tr>
<tr>
<td><strong>PFS</strong></td>
<td>Time from randomization until objective tumor progression or death</td>
<td>Quicker trial completion and smaller sample size when compared with OS, not affected by crossover or subsequent lines of treatment, appropriate for both cytotoxic and cytostatic</td>
<td>Declaring progression is subjective and susceptible to bias, difficult to correlate with clinical benefit and does not always translate to survival benefit</td>
</tr>
<tr>
<td><strong>TTP</strong></td>
<td>Time from randomization until objective tumor progression, not including death</td>
<td>Same as PFS</td>
<td>Same as PFS</td>
</tr>
<tr>
<td><strong>TTF</strong></td>
<td>Time from randomization to treatment discontinuation</td>
<td>Takes into account both efficacy and tolerability of drug</td>
<td>Affected by multiple subjective factors that can lead to treatment discontinuation such as patient and clinician preference</td>
</tr>
<tr>
<td><strong>CBR/DCR</strong></td>
<td>Percentage of patients achieving CR, PR and SD</td>
<td>Rapid assessment of antitumor activity, not affected by crossover, can capture disease stabilization with cytostatic drugs</td>
<td>Declaring progression is subjective and susceptible to bias</td>
</tr>
<tr>
<td><strong>ORR</strong></td>
<td>Percentage of patients achieving CR or PR</td>
<td>Rapid assessment of antitumor activity, not affected by crossover</td>
<td>Declaring progression is subjective and susceptible to bias, not suitable for cytostatic agents and low-grade indolent cancer types</td>
</tr>
</tbody>
</table>

**Table 1. Commonly used definitions, advantages and disadvantages of commonly used trial end points.**

CBR: Clinical benefit rate; CR: Complete response; DCR: Disease control rate; OS: Overall survival; ORR: Objective response rate; PFS: Progression-free survival; PR: Partial response; SD: Stable disease; TTP: Time to progression; TTF: Time to failure.
with an improvement in patient-reported outcomes (PROs). The term PROs has previously been defined as a ‘measurement of any aspect of a patient’s health status that comes directly from the patient (i.e., without the interpretation of the patient’s responses by a physician or anyone else)” [42]. PROs include health-related quality of life (HRQoL) which is a multidimensional patient-defined measure that accounts for the physical, emotional and social well-being of patients. Integrating measurement of PROs within randomized controlled trials (RCTs) is increasingly encouraged. Any improvement in PROs would complement any benefits seen in tumor-based end points and can therefore provide a stronger rational for the approval of new, potentially costly agents that may otherwise only offer a marginal PFS or OS benefit. Alternatively, PROs may indicate that the toxicity of an agent is negatively impacting on HRQoL. In this case, any PFS benefit, particularly in the absence of confirmed survival benefit, may be considered meaningless and prompt re-evaluation of whether approval of such a drug is justified.

PROs are increasingly being included within RCTs in oncology although the quality of reporting remains poor [43]. This is demonstrated by a recent review that included all Phase III RCTs published between 2007 and 2011. Forty-eight percent of trials reported PROs but the majority did not conform to the suggested recommendations in the CONSORT PRO extension guidelines [44]. The credibility of PROs in the current literature is limited by numerous methodological challenges such as determining the appropriate time and frequency that data should be captured, methods of dealing with missing data, ensuring adequacy and validity of existing tools and utilizing adequate statistical analysis and reporting methods. This is perhaps one of the reasons why all clinicians have not fully endorsed PROs as an end point in clinical trials. However, the development of the CONSORT PRO extension guidelines may help improve the quality of reporting and in turn increase the value that clinicians place on PROs. Interestingly, emerging evidence from a recent study suggests that HRQoL may have a role beyond the assessment of a patient’s welfare, as specific domains within HRQoL questionnaires were found to be of prognostic significance for different cancer sites [45].

Pancreatic NETs
Gastro-entero-pancreatic neuroendocrine tumors (GEP-NETs) are rare tumors with an estimated incidence of 5.25 per 100,000 per year [46]. GEP-NETs encompass a heterogeneous group that is subclassified by the site of origin, grade/proliferative tendency, level of differentiation and functionality, which refers to whether the tumor secretes biologically active peptides or hormones. Prognosis is variable and dependent on these characteristics and also on whether the tumor is resectable with curative intent or advanced. The WHO classification recognizes the inherent variability in behavior among GEP-NETs and differentiates between well-differentiated (low or intermediate grade) and poorly differentiated (high grade) tumors [47]. pNETs have a worldwide incidence of 0.2–0.4 per 100,000 per year and account for 12.1% of all GEP-NETs and 1–2% of all pancreatic tumors [48–50]. Prognosis of pNETs is influenced by both tumor grade and stage. A study reported 5 year survival rates of 75%, 62% and 7% for low-, intermediate- and high-grade pNETs, respectively. Five-year OS rates were demonstrated to be 92, 84, 81 and 57% using the American Joint Committee on Cancer and 100, 88, 85 and 57 using the European Neuroendocrine Tumor Society staging system for stage I, II, III and IV tumors, respectively [51].

Designing clinical trials and choosing appropriate end points for such a rare and heterogeneous condition,
such as the NET tumor group, which manifests with diverse clinical presentations that range from large volume, functional tumors with heavy symptom burden to asymptomatic low volume, low-grade tumors is very challenging. The prospect of poor trial accrual needs to be carefully balanced against the risk of including too heterogeneous a population which would undoubtedly impact on the clinical applicability of the trial results.

Considerations that need to be taken into account when choosing trial end points for Phase II trials in rare cancers

In early-phase trials, single-arm trials are often employed, as this requires a smaller sample size which is seen as an advantage in the trial design for rare cancers. However, the interpretation of single-arm trials relies heavily on historic controls, which in the case of NETs is limited by the diverse populations enrolled and the older classification systems used in previous trials. A recent review that included trials for NET that were undertaken between 2000 and 2011 highlighted some of the key methodological limitations of older NET trials. Out of the 46 articles evaluated, 92% were single-arm studies with variable sample sizes (range: 17–150) and heterogeneous populations. Moreover, in 28% of the trials the primary end point was not clearly defined and in some other trials, intention to treat analysis was missing, thereby limiting the validity of their results.

More recently, two new targeted agents, everolimus and sunitinib, have been licensed for use in advanced NETs. Herein we discuss the design of the two recent Phase II trials that investigated the antitumor activity of sunitinib and everolimus in the metastatic pNET setting, which subsequently led to the two agents being taken forward to Phase III trials. In 2008, Kulke and colleagues used an open-label, nonrandomized two-cohort, Simon two-stage design to test activity of sunitinib in advanced pNET and carcinoid tumors. The null hypothesis used suggested that the true ORR, which was the primary end point, was ≤5% versus the alternative hypothesis that the true response rate was ≥15%. They utilized a study power of 85% and a significance level of 5%. ORR in metastatic pNET patients was found to be 16.7%. Among the carcinoid patients, the ORR was 2.4%. Median TTP was 7.7 months in the pNET subgroup and 10.2 months in patients with advanced carcinoid. When compared with baseline, there were no significant differences in patient-reported quality of life during treatment. Based on these results it was concluded that sunitinib has antitumor activity in pNET, whereas its activity against carcinoid tumors could not be definitively determined.

In 2010, Yao and colleagues utilized a similar methodological design. They conducted an open-label, nonrandomized, Phase II study to assess the clinical activity of everolimus in patients with metastatic pNETs who experienced progression on or after chemotherapy. Patients were stratified by ongoing octreotide therapy at study entry. Similar to the study by Kulke and colleagues, a two-stage Simon design which allowed for early trial termination was used. The primary end point was ORR. A null hypothesis of an ORR less than 3% versus an alternative hypothesis of ≥10% ORR (80% power) was utilized. Based on a partial response (PR) rate of 9.6%, a stable disease (SD) rate of 67.8% and a median PFS of 9.7 months for patients treated with everolimus alone and a PR of 4.4%, an SD of 80% and a median PFS of 16.7 months in the combination group, it was concluded that daily everolimus, with or without concomitant octreotide long acting release (LAR), demonstrated antitumor activity in pNET.

These methodologically similar trials were part of the successful development of two targeted therapies for the treatment of advanced pNET, thereby suggesting that an open-label, single-arm design with ORR as the primary end point could be considered as an adequate design for the purpose of Phase II studies in rare cancers, including tumor types, such as NETs, that have a long median survival. Yao and colleagues demonstrated the importance of accurately defining the trial population and how stratification can be successfully utilized to control for the concomitant use of established interventions, such as octreotide. It would be unrealistic to expect large-scale prospective Phase II studies for each molecular subtype of NET. Nevertheless, as indicated by the trial by Kulke and colleagues, by clearly defining the population subgroups (e.g., pNET vs carcinoid) and by conducting retrospective analysis for each subgroup, differential activity for each molecular subtype can be adequately demonstrated. In addition, the inclusion of prognostic and predictive biomarker analysis, as exemplified by the analysis of Chromogranin A and neuron-specific enolase by Yao and colleagues, can further assist in defining the patients that are more likely to benefit from the experimental agent under investigation.

In the study by Kulke and colleagues, sunitinib failed to demonstrate a significant ORR in patients with advanced carcinoid and therefore in the subsequent Phase III trial, carcinoid patients were excluded. The TTP in the carcinoid subgroup was 10.2 months compared with 7.7 months in the pNET subgroup, with the pNET group achieving a higher ORR. In the absence of both a control group and a reliable historical control to refer to, it was difficult to evaluate whether the PFS magnitude that was seen in the carcinoid subgroup was of any clinical significance. This demonstrated one of
the greatest limitations of open-label, single-arm trials that utilize ORR as the primary end point. This is of even more relevance in cancer types with an indolent natural history, such as carcinoid, where response rates are expected to be low. Therefore, it can be argued that randomized controlled Phase II trials with PFS as the primary end point would be preferable in rare cancers such as pNET that are expected to have a low response rate and long postprogression median survival. This is of even greater significance in trials investigating cytostatic agents, such as everolimus, where disease stability is the expected clinical benefit and therefore time-based end points such as PFS would be more appropriate. The main concern at the design stage would be the feasibility with regard to recruitment rates within a reasonable time frame. However, as discussed below, multicenter international trials can facilitate trials with larger sample sizes. Additionally, compensatory pragmatic adjustments in the statistical considerations, such as the acceptance of a greater type I error, can also aid in reducing the sample size requirements and thereby allowing for a more feasible Phase II RCT.

Considerations that need to be taken into account when choosing trial end points for Phase III RCTs in rare cancers

Table 3 summarizes the characteristics of the key Phase III RCTs for the treatment of pNET. This table demonstrates the recent shift in the choice of primary end points from response rate which was preferred in studies which investigated the role of cytotoxic chemotherapy to PFS which appears to be the primary end point of choice for trials that investigate the role of newer cytostatic targeted therapies.

Larger trials in rare cancers, such as pNET, are feasible provided that national and international collaborations are formed. This is clearly highlighted by RADIANT-3 which was a double-blind placebo-controlled RCT with a sample size of over 400 [58]. They enrolled patients with advanced-, low- or intermediate-grade pNET with radiological progression over the previous 12 months and compared everolimus with best supportive care. The median PFS according to central assessment was 11.4 months for the everolimus group, compared with 5.4 months for the placebo group (hazard ratio [HR]: 0.34; p < 0.001). Median OS was not reached at the time of this analysis, and no significant difference between the groups was observed (HR for death with everolimus: 1.05; p = 0.59).

Similarly, in the Phase III RCT that investigated the role of sunitinib in advanced pNET, 171 patients received either sunitinib (37.5 mg) or placebo once daily [57]. Based on early results that favored the sunitinib arm, the independent data monitoring committee recommended trial termination earlier than the prespecified interim analysis. Median PFS was 11.4 months in the sunitinib group compared with 5.5 months in the placebo group (HR: 0.42; p < 0.001). The initial OS analysis, based on 18% of events, suggested an OS benefit for the sunitinib group (HR for death was 0.41 [95% CI: 0.19–0.89; p = 0.02]). However, further analysis based on 43% of the events did not show a statistically significant difference between the two groups (HR: 0.74 [0.47–1.17]) [60]. It has been suggested that the early unplanned efficacy assessment may have resulted in an overestimation of the PFS magnitude. The FDA recommended that such practice should be discouraged as unplanned early analyses, with subjective end points such as PFS, can increase the risk of identifying an erroneous false positive result, thereby leading to a type I error and overestimating the treatment effect of the experimental agent [60].

The use of PFS over OS as a primary end point is justified by the fact that pNET is characterized by a long natural history and therefore a meaningful change in OS would require both a large sample size and a long follow-up, thereby rendering such a trial unfeasible. This was supported by the consensus report of the National Cancer Institute NET clinical trials planning meeting, where it was recommended that PFS is the preferred primary end point [60]. Nevertheless, an intervention that leads to a sizable PFS gain might be expected to show at least a numerical, if not a statistically significant benefit in OS. However, this was not the case in both the everolimus and sunitinib trials. It is plausible that in both RADIANT-3 and the sunitinib trial the OS data were significantly confounded by the crossover that took place in both trials, with estimated crossover rates being 73% and 69% for RADIANT-3 and sunitinib trial, respectively [57,58]. The effect of crossover design on OS has been a matter of considerable controversy. Confounding the interpretation of OS inadvertently increases the significance placed upon the tumor-based end points such as PFS and ORR, which in itself is a contentious issue, as it remains unclear whether or not extending PFS provides a discernible clinical benefit. On the other hand, the ethical considerations of denying a potential treatment option to patients that have limited lines of approved treatments needs to be considered, if crossover were to be prohibited. Moreover, omission of crossover is likely to have a negative impact on accrual, as patients are more likely to support a placebo-controlled trial that allows crossover.

The importance of assessing HRQoL in rare cancers

While the use of PFS as a primary end point is not without limitations, it has become a widely accepted end
Clinical trial end points relevant to patients & society for rare cancers

**Clinical Trial Methodology**

Point for Phase III trials to hasten drug approval for rare cancers where there is a clinical unmet need. Although knowledge of whether a particular agent extends life is of paramount importance to most patients, in cases where capturing survival data are not always possible, such as with rare cancers, patients may be satisfied with just the knowledge that a particular intervention reduces symptoms and consequently improves HRQoL. Interestingly, there appears to be an association between PFS and HRQoL in some tumor types (see Table 4) [62]. However, to our knowledge, only five studies have investigated this [63]. The limited data available may be related to publication bias where negative studies have not been published. A recent study of Afatinib in non-small-cell lung cancer (NSCLC), which is a highly symptomatic malignancy, evaluated the relationship between PFS and HRQoL [64]. The study did illustrate that at the time of progression, patient’s HRQoL was worse. It is uncertain whether the association between PFS and HRQoL in NSCLC where patients are highly symptomatic would hold true for patients with pNET who tend to be less symptomatic.

Two recent RCTs in the field of pNETs evaluated HRQoL as a secondary end point. Both trials util-

---

**Table 3. End points used in selected Phase III studies of pancreatic neuroendocrine tumors.**

<table>
<thead>
<tr>
<th>Study (year)</th>
<th>n</th>
<th>Study location</th>
<th>Population</th>
<th>Comparators</th>
<th>Primary end point</th>
<th>Secondary end points</th>
<th>HRQoL assessed?</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moertel et al. (1980)</td>
<td>84</td>
<td>USA</td>
<td>Unresectable or metastatic islet cell carcinoma</td>
<td>Streptozocin vs streptozocin + fluorouracil</td>
<td>Uncertain but ORR, CR, median survival and safety were reported</td>
<td>As per primary end point</td>
<td>No</td>
<td>[55]</td>
</tr>
<tr>
<td>Moertel et al. (1992)</td>
<td>105</td>
<td>USA</td>
<td>Unresectable or metastatic islet cell carcinoma</td>
<td>Chlorozotocin vs streptozocin + doxorubicin vs streptozocin + fluorouracil</td>
<td>Rate of tumor regression (defined broadly as either reduction in perpendicular diameters by 50%, reduction in size of hepatomegaly by 30% and hormonal response in patients without measurable disease)</td>
<td>TTP median survival</td>
<td>No</td>
<td>[56]</td>
</tr>
<tr>
<td>Raymond et al. (2011)</td>
<td>171</td>
<td>42 centers in 11 countries</td>
<td>Advanced well-differentiated pancreatic NET with documented disease progression with 12 months from baseline</td>
<td>Sunitinib vs placebo</td>
<td>PFS</td>
<td>ORR, OS, Safety</td>
<td>Yes</td>
<td>EORTC QLQ-C30</td>
</tr>
<tr>
<td>Yao et al. (2011)</td>
<td>410</td>
<td>International</td>
<td>Advanced well-differentiated pancreatic NET with documented disease progression with 12 months from baseline</td>
<td>Everolimus vs placebo</td>
<td>PFS</td>
<td>ORR, OS, Safety</td>
<td>No</td>
<td>[58]</td>
</tr>
<tr>
<td>Meyer et al. (2014)</td>
<td>86</td>
<td>23 UK centers</td>
<td>Chemo-naive metastatic or unresectable NETs of pancreatic (48%), gastrointestinal foregut (20%) or unknown primary site (33%)</td>
<td>Capecitabine + streptozocin vs capecitabine + streptozocin + cisplatin</td>
<td>ORR, Biochemical response</td>
<td>Safety, PFS, OS</td>
<td>Yes</td>
<td>EORTC QLQ-C30</td>
</tr>
</tbody>
</table>

CR: Complete response; EORTC QLQ-C30: European Organisation for the Research and Treatment of Cancer Quality of Life Questionnaire Core 30; OS: Overall survival; ORR: Objective response rate; PFS: Progression-free survival; TTP: Time to progression.
lized the European Organisation for the Research and Treatment of Cancer Quality of Life Questionnaire Core 30 (EORTC QLQ-C30), which together with the functional assessment of cancer therapy-general are the most widely used tools for the assessment of HRQoL in oncology trials [69].

In the RCT that investigated the role of sunitinib in pNET, HRQoL results were only briefly discussed in the main publication and the data were not graphically presented. The questionnaire completion rate was 84% and despite the improved PFS seen in the sunitinib group, there was no overall difference in global HRQoL between the sunitinib and placebo groups.

The RCT by Raymond and colleagues compared two chemotherapy regimens: capecitabine and streptozocin vs capecitabine, streptozosin and cisplatin for the treatment of NETs [57]. In total, 48% of the trial population had a diagnosis of pNET. There was no significant difference with regard to the primary end point, ORR. HRQoL was a secondary end point. In this study, the EORTC QLQ-C30 questionnaire completion rate was poor, with 86% completion rate at baseline, 50% after three cycles of treatment and 43% after completing 6 months of treatment. HRQoL results were presented using box plots of global health status score. Patients that received the triplet regimen had significant deterioration in their global health status after three cycles of treatment (>15 points reduction, p= 0.05) with further deterioration after 6 months. There was a smaller degree of deterioration in the global health status score in the group of patients that received the doublet regimen.

It is encouraging to see that recent RCTs in rare tumors, such as pNETs, included HRQoL as a secondary end point. However, these RCTs also exemplify some of the limitations and challenges that exist in the field of PROs for the clinical trials of rare tumors. In addition to limitations that apply in all tumor types, such as poor completion rates and poor outcome

<table>
<thead>
<tr>
<th>Tumor type</th>
<th>Anticancer therapy utilized</th>
<th>Name of HRQoL assessment tool(s) used</th>
<th>Is there any association between PFS and HRQoL?†</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colorectal cancer</td>
<td>Panitumumab + BSC vs BSC</td>
<td>EORTC QLQ-C30, EQ-SD, EQ-VAS, NCCN FCSI to measure symptomatology</td>
<td>Lack of disease progression is associated with better symptom control, HRQoL and survival</td>
<td>[65]</td>
</tr>
<tr>
<td>Breast cancer</td>
<td>Lapatinib + capecitabine vs capecitabine alone</td>
<td>FACT-B, EQ-SD, EQ-VAS, FACT-G, TOI</td>
<td>Patients with an objective tumor response or stable disease showed a clinically meaningful differences in QOL compared with patients with progressive disease</td>
<td>[66,67]</td>
</tr>
<tr>
<td>Renal Cell Cancer</td>
<td>Pazopanib vs placebo</td>
<td>EORTC QLQ-C30, EQ-SD, EQ-VAS</td>
<td>Patients experiencing tumor response/stabilization may also have better HRQoL compared with those without this response</td>
<td>[68]</td>
</tr>
<tr>
<td>NSCLC</td>
<td>Afatinib + BSC in LUX-Lung 1 study, OR Afatinib vs Cisplatin/Pemetrexed in Lux-Lung 3 study</td>
<td>EORTC QLQ-C30, EQ-SD, EQ-VAS</td>
<td>Tumor progression associated with statistically significant worsening in HRQoL</td>
<td>[64]</td>
</tr>
</tbody>
</table>

†Conclusion reached by the referenced study authors.

EQ-SD: Overall utility and EQ-visual analogue scale (VAS) are components of EuroQOL disease–generic questionnaire and assess health status.

reporting in the main manuscript publications, rare tumors are also limited by the lack of or slow development of validated tumor specific HRQoL tools. Both of the aforementioned pNET RCTs utilized the generic EORTC QLQ–C30 questionnaire, as at the time that these RCTs were designed, a validated NET specific HRQoL module was not available. In contrast to this, in common tumor types, tumor-specific modules that supplement the core HRQol questionnaires have been used for over 10 years [66–67,70–72]. Patients with NETs often present with distinct clinical syndromes that can impact on their HRQoL, particularly if their tumors are functional. Moreover, qualitative analyses also indicate that during the course of their treatment, patients with NETs are also faced with challenges relating to the often inferior service provision structure that exists for patients with such rare tumors [73]. More recently, the acknowledgement of the distinct clinical course that patients with NETs experience led to the development and validation of QLQ-GINET21, a tool that was specifically designed for the assessment of HRQoL in the gut, pancreas and liver NETs [74]. It is hoped that the future use of this tool in the field of NET research will facilitate better quality research in the field of PROs for NETs. It would be unrealistic to expect distinct HRQoL tools to be designed for each subtype of NET and as a result of this the QLQ-GINET21 questionnaire has been validated in most common NET subtypes including pNETs [75].

As the cost of new drugs continues to rise, demonstration of greater efficacy benefits are likely to be required for drugs to be licensed. However, for drugs with a similar efficacy outcome, society may be more willing to accept expensive new drugs for rare cancers compared with common cancers. This is perhaps best highlighted by the Phase III trials of the drug regorafenib in colorectal cancer and GIST. In the GRID trial conducted in patients with advanced GIST who had failed treatment with imatinib or sunitinib due to poor tolerance or progression, regorafenib was associated with a longer median PFS compared with placebo (4.8 vs 0.9 months; HR: 0.27; p < 0.0001) but no difference in OS, although the trial permitted crossover [76]. In the CORRECT trial, patients who had received all standard approved treatment for colorectal cancer and had evidence of progression or intolerance within 3 months of their last dose of standard therapy were eligible [77]. The trial showed an OS benefit in favor of regorafenib compared with placebo (6.4 vs 5 months; HR: 0.77; p = 0.0052). Additionally, PFS was improved with regorafenib (1.9 vs 1.7 months; HR: 0.49, p < 0.0001). However, regorafenib is not currently available via the National Health Service for patients in the UK or the national cancer drugs fund for patients in England, whereas up until recently, it could be accessed via the national cancer drugs fund for advanced GIST. The incidence of colorectal cancer is 470 per million compared with the incidence of approximately 10 per million in the case of GIST. This suggests that society may be more willing to pay for expensive treatments for rare cancers rather than common cancers as overall they represent less of a financial burden.

Are all cancers actually rare cancers?

Over the years, improvements in genomic technologies have led to the recognition that common cancers encompass several molecular entities which behave and respond differently to therapeutic agents. This has led to the reclassification of many cancers by their molecular profile, in addition to their site of origin and histological subtype. Trial designs have frequently been altered to include enriched populations with a specific molecular aberration that is likely to respond to the anticancer agent being investigated. Consequently, trials of therapeutic agents for common cancers are transforming into smaller trials for molecular subtypes with incidence rates that are similar to rare cancers. Therefore, lessons learned from trials of rare cancers will become increasingly important and similarly, any successes from trial design for rare molecular subtypes of common cancers may be incorporated into studies for rare cancers in the future.

The presence of ALK rearrangements in lung cancer is a clear example of a rare molecular subtype of a common tumor. NSCLC has an estimated worldwide incidence of 1.3 million and ALK rearrangements occur in 5% of NSCLC cases [78] which will translate into an incidence of <6 per 100,000 per year which has previously been used to define rare cancers. Crizotinib is a small molecule tyrosine kinase inhibitor targeting ALK, MET and ROS1. The FDA granted accelerated approval for crizotinib in ALK positive NSCLC based on durable ORRs of 53% and 61% in two single-arm trials [79,80]. Regular approval was then granted based on a randomized, multinational, open-label trial enrolling 347 ALK positive patients with metastatic NSCLC after screening 4967 patients [81]. The primary end point was PFS and the study demonstrated an improvement in both PFS and ORR in favor of crizotinib compared with chemotherapy in ALK positive NSCLC patients who had disease progression after platinum-based doublet chemotherapy. No difference in OS was evident in a planned interim analysis. The accelerated approval granted on the basis of ORR seen in single-arm studies parallels the situation seen in studies of rare cancers. There are increasing efforts to ensure that drugs are speedily accessible to patients with rare cancers to
prevent them from being disadvantaged by the sheer rarity of their tumor and the time that a randomized Phase III trial may take to accrue. However, accelerated approvals may compromise long-term toxicity for these patients. The focus should therefore be a collaborative effort worldwide, which will allow a larger proportion of patients with a specific rare tumor to be identified and help quicken and increase accrual into larger studies and hopefully lead to more meaningful results.

**Conclusion & future perspective**

We expect that in the next five to ten years, research for rare cancers will improve with the establishment and expansion of international collaborations such as the International Rare Cancers Initiative. The ongoing recognition of the specific challenges faced by patients with rare cancers will prompt restructuring of services which will hopefully lead to centers with greater expertise for specific rare cancers. This is in turn will

---

**Executive summary**

**A discussion of the strengths & limitations of end points used in oncology trials**

- Overall survival (OS) remains the gold standard primary end point but its use is limited by the need for a larger sample size, longer follow-up and subsequent increased cost.
- In rare cancers the use of OS as a primary end point would result in sample sizes that would be unfeasible to attain within a reasonable time frame, thereby hindering new drug development in a much needed clinical field.
- Surrogate end points such as progression-free survival (PFS) are increasingly used as primary end points. However, they do not always translate into survival gains and as such their direct clinical benefit for patients is difficult to discern.
- The use of PFS in unblinded trials is subject to bias due to the inherent subjective element associated with declaring progression. Blinded independent central review can minimize this but increases the potential for measurement variability, the cost and complexity of the trial and should therefore be reserved for smaller trials or alternatively, a sample-based blinded independent central review might be employed.
- The magnitude of PFS benefit is subject to overestimation as the exact progression date could have occurred prior to the latest scan. Specific interval censoring statistical methods should therefore be utilized to account for this.
- The choice of tumor response criteria should be carefully selected as this would impact on all tumor-centered end points. The nature of the cancer type and treatment (cytotoxic vs cytostatic vs immunotherapy) should influence decision-making.

**Considerations that need to be taken into account when choosing trial end points for Phase II & Phase III trials in rare cancers**

- In rare and heterogeneous cancers, single-arm trials that use surrogate tumor-centered end points are difficult to interpret due to the lack of reliable historical controls. Therefore, randomized controlled trials (RCTs) are preferred.
- International collaborations are essential for the successful and timely accrual in rare cancer RCTs.
- Compensatory statistical adjustments such as the acceptance of a greater type 1 error may be required in order to allow for a more feasible sample size in rare cancer RCTs.
- Patient-reported outcomes (PROs) encompassing health-related quality of life (HRQoL) questionnaires are increasingly used in RCTs of common and rare cancers.
- Any improvement in PROs would complement any benefits seen in tumor-based end points and can therefore provide a stronger rational for the approval of new, potentially costly agents that may otherwise only offer a marginal PFS or OS benefit.

**The importance of assessing health-related quality of life in rare cancers**

- The credibility of PROs in the current literature is limited by numerous methodological challenges such as: determining the appropriate time and frequency that data should be captured, methods of dealing with missing data, ensuring adequacy and validity of existing tools and utilizing adequate statistical analysis and reporting methods.
- The use of cancer specific HRQoL modules should be used in addition to the generic questionnaires. The creation and validation of these tools has been slower to develop in rare cancers.

**Are all cancers actually rare cancers?**

- The reclassification of many common cancers by their molecular profile, in addition to their site of origin and histological subtype, has led to smaller trials that encompass populations with similar incidence rates to that seen in rare cancers.
- Lessons learned from trials of rare cancers will become increasingly important and similarly, any successes from trial design for rare molecular subtypes of common cancers may be incorporated into studies for rare cancers in the future.
facilitate the prioritization of areas for research and will allow for the development of novel trial methodologies. We expect that future trials will use a combination of tumor-centered end points, such as PFS along with safety reporting and PROs in order to demonstrate direct clinical benefit for patients with rare cancers. As the quality of collecting, analyzing and reporting PROs improves, the integration of PROs within trials may become a compulsory requirement for regulatory bodies to approve a drug. As public expectations and the cost of drugs continue to rise, funding bodies are likely to demand greater efficacy benefits for new drugs. However, society may be more willing to pay for expensive treatments for rare cancers rather than common cancers as overall they represent less of a financial burden.

References
Papers of special note have been highlighted as:
• of interest; •• of considerable interest


•• Provides a definition of rare cancer and estimates the rare cancer burden in Europe.


• Commentary discussing the unique challenges faced by patients with rare cancers and the development of the International rare cancers initiative in an attempt to combat some of these.


• A study examining the different endpoints used to support regular and accelerated approval of cancer drugs between 2002 and 2012.


16 Krajewski KM, Guo M, Van Den Abbeele AD et al. Comparison of four early posttherapy imaging changes (EPTIC; RECIST 1.0, tumor shrinkage, computed tomography tumor density, Choi criteria) in assessing outcome to vascular endothelial growth factor-targeted
therapy in patients with advanced renal cell carcinoma. 


26 Saad ED, Katz A, Hoff PM, Buyse M. Progression-free survival as surrogate and as true end point: insights from the breast and colorectal cancer literature. *Ann. Oncol.* 21(1), 7–12 (2010).


39 Foster NR, Qi Y, Shi Q et al. Tumor response and progression-free survival as potential surrogate end points for overall survival in extensive stage small-cell lung cancer: findings on the basis of North Central Cancer Treatment Group trials. *Cancer* 117(6), 1262–1271 (2011).


• Outlines how the CONSORT patient-reported outcome extension aims to improve reporting of patient-reported outcomes in clinical trials.
Clinical trial end points relevant to patients & society for rare cancers

Clinical Trial Methodology


• A succinct review of recent trials that investigated the role of targeted therapies for the treatment of neuroendocrine tumor.


** A comprehensive review outlining the advantages and disadvantages of progression-free survival and other end points in cancer trials


