Rare cancers 3

Research methods to change clinical practice for patients with rare cancers

Lucinda Billingham, Kinga Malottki, Neil Steven

Rare cancers are a growing group as a result of reclassification of common cancers by molecular markers. There is therefore an increasing need to identify methods to assess interventions that are sufficiently robust to potentially affect clinical practice in this setting. Methods advocated for clinical trials in rare diseases are not necessarily applicable in rare cancers. This Series paper describes research methods that are relevant for rare cancers in relation to the range of incidence levels. Strategies that maximise recruitment, minimise sample size, or maximise the usefulness of the evidence could enable the application of conventional clinical trial design to rare cancer populations. Alternative designs that address specific challenges for rare cancers with the aim of potentially changing clinical practice include Bayesian designs, uncontrolled n-of-1 trials, and umbrella and basket trials. Pragmatic solutions must be sought to enable some level of evidence-based health care for patients with rare cancers.

Lancet Oncol 2016; 17: e70–80

This is the third in a Series of three papers about rare cancers

Cancer Research UK Clinical Trials Unit, University of Birmingham, Birmingham, UK (Prof L Billingham PhD, K Malottki MA, N Steven PhD); **and University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK** (N Steven)

Correspondence to:
Prof Lucinda Billingham, Cancer Research UK Clinical Trials Unit, University of Birmingham, Vincent Drive, Edgbaston, Birmingham B15 2TT, UK
L.J.Billingham@bham.ac.uk

Introduction

There are no fixed criteria for the definition of a rare cancer, but a malignant disorder that has an incidence of six or less per 100 000 people per year is commonly classed as rare.[1] Investigation of treatments for rare cancers is important—collectively more than 20% of all cancers diagnosed are rare, and ethically patients with a rare cancer have as much entitlement to evidence-based health care as those with common cancers.

Rare cancers were historically defined by their characteristic natural history and histopathological appearances. Increasingly, rare subtypes of common cancers are defined on the basis of molecular markers that have a specific natural history or potentially a particular response to treatment. Furthermore, subtypes can be defined by biomarkers that directly relate to key oncogenic pathways targeted by drugs or other interventions, giving rise to testable therapeutic hypotheses.[2] Deep genome sequencing of cancers is increasingly revealing a few common driver aberrations but a high number of cancer genes mutated at frequencies of less than 5%.[3]

Increasing stratification of cancers and personalisation of treatment justifies an increasing need to identify methods for assessment of interventions that are sufficiently robust to change clinical practice in this setting. A repertoire of research methods is needed to recognise that a cancer defined as rare covers a range of incidences from the more common to extremely scarce and different methods might be appropriate at different points on the range.

The aim of our Series paper is to identify methods for the investigation of interventions for rare cancers that can affect clinical practice—an arena dominated by large phase 3 randomised controlled trials in which experimental treatments are compared with standard of care. The methods we discuss retain the core principles of clinical trials to provide unbiased evidence about the effect of interventions but are specifically relevant to small populations.

Challenges of conventional trials in rare cancer

The conventional phase 3 clinical trial design for affecting clinical practice is a hypothesis-testing randomised controlled trial with parallel-group treatment comparisons. Such trials are at the highest level of the hierarchy of evidence for assessing the effects of interventions on diseases. In this approach, frequentist statistical analysis is typically used to test the null hypothesis that there is no treatment effect against the alternative hypothesis that there is a treatment effect of a prespecified clinically meaningful size. In phase 3 cancer trials, the most common primary outcome measure for treatment comparison is survival time with the hazard ratio (HR) as a measure of treatment effect and statistical design based around a required number of events rather than patients. The minimum clinically relevant treatment effect size that is deemed sufficient to change clinical practice will typically be moderate in size and is a key driver for the sample size. Trials are designed such that the chance is small of incorrectly rejecting the null hypothesis (ie, a false positive conclusion, or type I error) by setting the significance level to 5%, and such that there is a good chance (often 90%) of rejecting the null hypothesis (at a 5% significance level) when the prespecified minimum clinically relevant treatment effect truly exists (ie, power, which is the probability of a true positive conclusion, the complement of a type II error).

The difficulty in the context of rare cancers is that conventional trial design usually demands an unfeasibly large number of patients and small trials would be underpowered and less likely to produce robust, conclusive results. For many rare cancers, the quality of existing evidence for all aspects of clinical management is clustered at the lowest level: retrospective case series and case reports, very few prospective studies, and a dearth of randomised trials. Thus, for rare cancer populations, the challenge is not

only recruitment of sufficient numbers of patients to clinical trials but also establishment of a standard of care against which to test new interventions.

Overview of methodological options

Several reviews of methods for clinical trials in rare diseases have been published, both general[4–11] and

	Applicability to rare cancers	Regulatory endorsement
Create national and international networks to maximise recruitment[4,20]	Can help with accrual to allow sufficient sample size for a conventional design Recognised strategy in rare cancer[21]	Accepted by funders in IRCI trials[22]
Maximise trial duration[4,20]	For time-to-event outcomes (eg, overall survival time), for which statistical power is determined by number of events, increasing duration can improve power (more events) in a small sample size	None
Select trial population to minimise sample size[4,20]	High-risk patients will have more events, so fewer patients needed Selection of patients likely to gain the most benefit will reduce the required sample size	Mentioned in FDA guidance[17]
Use relaxed statistical error rates; plan to incorporate underpowered studies into a meta-analysis[4,20]	Larger significance levels or lower power allows for reduced sample sizes Compromise to the robustness of the conclusions might be acceptable	Accepted by funders in IRCI trials[22]
Use external controls to reduce sample size[7,8,20]	Common for phase 2 cancer trials Comparison will be prone to bias, but if external data are robust, results could change clinical practice	Common design in FDA approvals;[19] mentioned in EMA and FDA guidance[16,17]
Select outcome measures to minimise sample size or maximise usefulness[4,13,20]	Composite or surrogate outcome can increase number of events or effect size and thereby reduce the sample size Continuous measures can have greater statistical power than an associated dichotomous variable and thus reduce sample size needed	Mentioned in EMA and FDA guidance[16,17]
Minimise noise to maximise usefulness[7]	Increase statistical efficiency by ensuring good-quality data and complete follow-up and by adjusting for discrepancies in baseline characteristics	Mentioned in EMA guidance[16]
Design and analyse trial in a Bayesian framework[4,6,7,13,20]	Limited available populations can be used in a randomised controlled trial to reduce uncertainty about treatment effect size	Accepted by funders in IRCI trials;[22] mentioned in EMA guidance[16]
Use observational studies to gather information about treatment outcomes[4,15,20]	Prospectively planned observational studies can be considered Could be directly linked to idea of serial uncontrolled n-of-1 trials	Mentioned in FDA guidance[17]
Use n-of-1 trials to collect treatment outcome data prospectively;[4,6–8,13,15,16] potentially combine a series in a random effects meta-analysis	Not generally applicable to the cancer setting because cancer is not a stable disease, but could be adapted to incorporate only periods of experimental treatment, which could be compared with predicted natural history with standard treatment (ie, uncontrolled n-of-1 trial) Superior to a case study because it is preplanned in a trial protocol	Mentioned in EMA guidance[16]

IRCI=International Rare Cancers Initiative. FDA=Food and Drug Administration. EMA=European Medicines Agency.

Table 1: **Summary of methods applicable to trials of rare cancer that could change clinical practice**

	Reason for being recommended for rare diseases	Reason for less applicability in rare cancer
Efficient designs	Several randomisations, which allows for several interventional questions (eg, factorial design,[4,8] ranking and selection designs,[6,8] multiarm multistage trials). One randomised controlled trial with several treatment groups would have smaller sample than several randomised controlled trials each with one experimental treatment group (sample size for individual question remains the same). Fewer patients needed for ranking and selection designs, but number of errors is higher	Such designs are used in cancer setting but the complexity might necessitate an unfeasible sample size for a rare cancer population
Response adaptive randomisation to reduce sample size[4,6,13,16]	Allows allocation ratio to change over the accrual period according to balance of observed successful outcomes while minimising overall sample size. Encourages enrolment: patients have greater chance of receiving the potentially better treatment, which can reduce overall sample size	Short-term outcome measures necessary to enable timely adaptation of randomisation ratio—not generally possible with time-to event outcomes. Could be applicable if short-term surrogate measures are deemed sufficient to affect clinical practice. Cannot guarantee reduction in sample size
Adaptive designs to recalculate sample size at interim analysis[6,14]	Interim data can be used to provide better estimates of required sample size, which could result in a reduction	Applicable to rare cancer but not guaranteed to reduce sample size and thus could be a risky approach
Sequential boundary designs	Repeatedly assess cumulative data in relation to prespecified stopping boundaries for efficacy and futility.[4,6,13,14,16] Potential to stop trial with a reduced sample size compared with fixed design	Not commonly used in cancer trials because survival-type outcome measures are often not observed quickly enough for this approach to be practical. Limited power for interim decisions
Crossover trial design to reduce sample size[4,6–8,13]	Each participant acts as their own control, thereby reducing variability in treatment effect and increasing statistical efficiency: fewer patients needed	Requires degree of disease stability such that if treatment is stopped, then disease will return to pre-treatment state, which is not typical behaviour for cancer. Cancer interventions generally aim to modify disease, which further invalidates approach

Table 2: **Clinical trial methods for rare disease that are less applicable in rare cancer**

population-specific[12–15] and in the form of regulatory guidelines.[16,17] Reviews of methods as applied to rare cancers include commentaries,[18] a specific review of trials underpinning approvals by the US Food and Drug Administration (FDA),[19] a European consensus position paper,[20] and a publication from the International Rare Cancers Initiative,[21] one of the leading organisations for driving research into rare cancers, which provides an overview of methods being implemented under their banner.[22]

The methods that are advocated in reviews relating to rare diseases that could affect or change clinical practice in the rare cancers setting are summarised in table 1. Some methods are less applicable in general to the cancer setting (table 2). Methods relevant for rare cancers are described in two broad sections in the paper: strategies that enable the application of conventional clinical trial design to rare cancer populations and alternative designs that address specific challenges for rare cancers with the aim of changing clinical practice.

The European Medicines Agency and FDA guidelines[16,17] acknowledge the challenge of trials in rare diseases and the need for flexibility from a regulatory perspective, and suggest that less conventional approaches might be acceptable, although they concede that deviation from the standard is uncommon and would need to be justified. Both sets of guidelines stress the importance of drug development being based on a sound understanding of the biology of the disease and the behaviour of the drug, and mention a range of specific approaches that should be considered (table 1). Compromises to the conventional design might result in trials that superficially resemble phase 2 designs, but such trials should still be branded as phase 3 if the aim is to affect clinical practice. No single trial design should be deemed acceptable by licensing and funding bodies for cancers defined as rare. Instead, the methods should be tailored both to the frequency of the cancer and to the detail of its clinical and biological characteristics (figure 1).

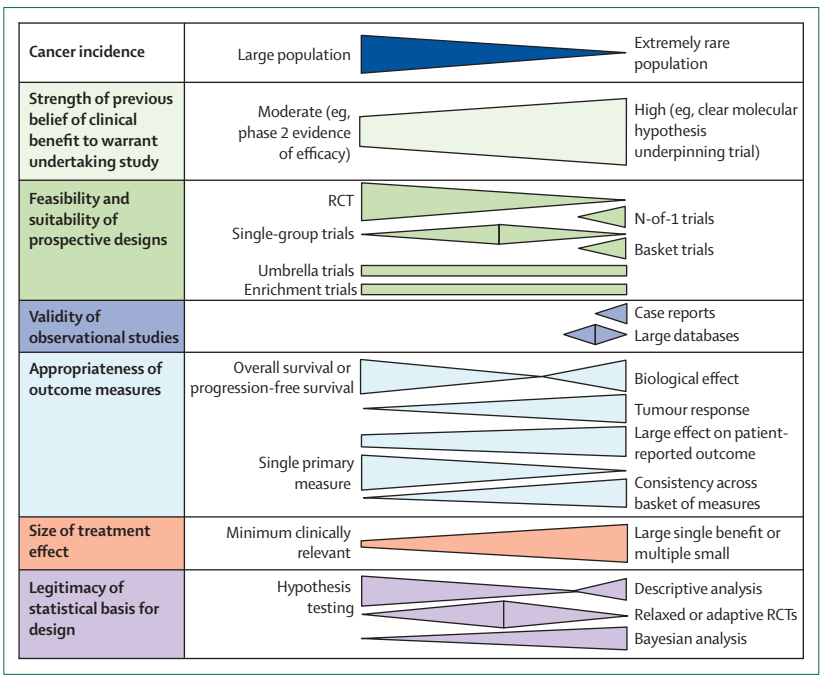

Figure 1: **Research methods to change clinical practice in relation to the range of cancer incidence**
This figure is not to scale. Cancer incidence is shown, from the conventional setting of common cancers with large populations on the left, to extremely rare cancers on the right by the taper of the blue triangle. The changes in importance and relevance of different notions and methods across the range of incidence are shown by the changing width of the shapes. All trials require prior belief in the efficacy of the experimental intervention, but a high level of prior evidence—eg, based on molecular selection of patients and targeting—might permit small trials for very rare cancers (pale green). By their nature, umbrella and enrichment trials are applicable to common cancers and might generate rare cancer populations. Conventional randomised trials are of decreasing relevance as cancer frequency decreases, n-of-1 and basket designs might be justified as practice-changing trials for rare but not commoner cancers, and single-group trials to change practice might be justified for less common cancers but be unfeasible for very rare populations (dark green). Observational data might change clinical practice for rare cancers but do not replace prospective controlled trials if such trials are feasible (middle blue). Survival and progression-free survival are gold standard outcomes in cancer trials, but other outcomes—eg, profound clinical response, improved wellbeing—might change practice in rare cancer trials in molecularly selected patients or if a basket of outcomes showed consistency patient by patient (pale blue and orange). Hypothesis testing remains the gold standard in clinical trials, but might be modified for less common cancers and unfeasible in very rare cancers, whereas a Bayesian probabilistic design has particular relevance for rare cancers (purple). RCT=randomised controlled trial.

Application of conventional designs to rare cancers

Maximise recruitment

Recruitment can be maximised by designing and running the trial as an international collaboration and by engaging full support from the specialist centres that accumulate the rare cancer population, with the aim of near 100% recruitment of eligible patients. This approach is long established in paediatric cancer trials and is key to the International Rare Cancers Initiative,[21] but is no trivial challenge. In the setting of low-quality evidence, clinical practice is probably diverse and opinions entrenched, making consensus on which questions to prioritise and the standards for comparative trials difficult to negotiate. Use of a more flexible, pragmatic design with different randomisation options might enable all partners to participate (eg, figure 2, panel 1), although the added complexity might further increase the sample size needed.

Recruitment can also be maximised by broadening eligibility criteria, which should be balanced against the possibility of dilution of the treatment effect. Trial conduct for international studies can be hampered by an inability to engage a single international sponsor, but this problem can be overcome by engaging multiple sponsors using the same trial protocol and identifying a single trial unit as a coordinating centre to combine data in a single analysis. If the logistic difficulties can be overcome, then recruitment from a broad global population with complete support from the clinical community might result in a sufficient sample size to implement an adequately powered randomised controlled trial.

Selection of population and treatment

One of the key drivers of sample size calculations is the size of the treatment effect that the trial is powered to detect. Designing a trial powered to detect a nominally large treatment effect is not useful for rare cancers

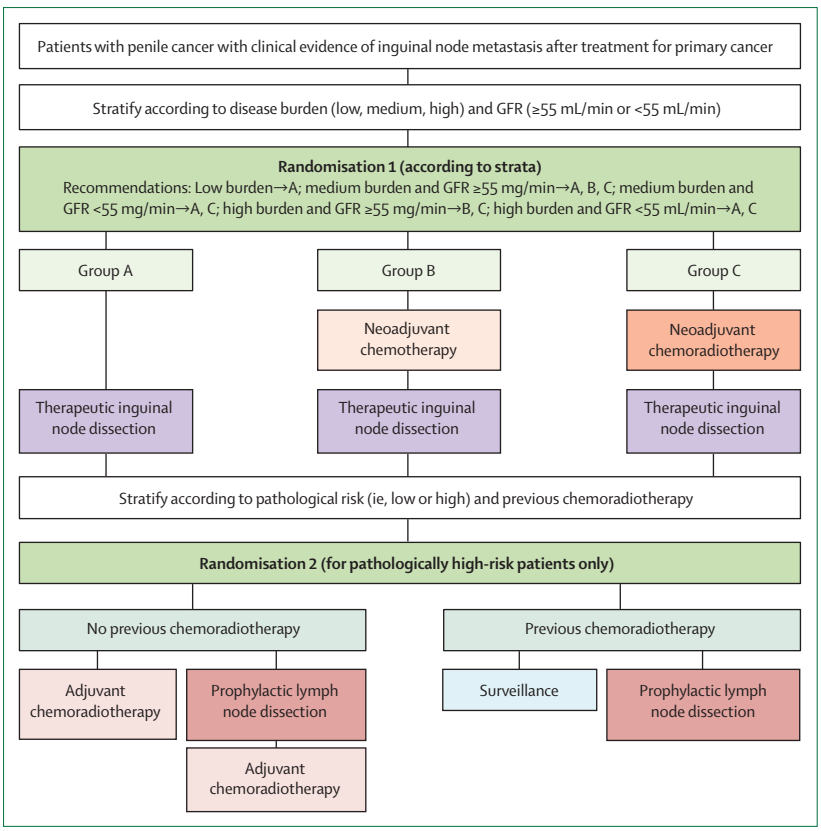

Figure 2: **Methods for a phase 3 rare cancer trial: the InPACT trial**
Trial registration: CRUK/13/005; IRCI 004; NCT02305654. GFR=glomerular filtration rate.

Panel 1: **Key features of phase 3 trial design for a rare cancer: the InPACT trial**

- Maximisation of recruitment through international collaboration managed by the International Rare Cancers Initiative
- Trial coordinated on international level by Institute for Cancer Research with country-specific sponsorship model (the Institute for Cancer Research is the UK sponsor)
- Stratification of disease with strata-specific randomisation options allows all patients to participate
- Several randomisations allow different questions along the patient pathway to be addressed
- Single, clinically relevant primary outcome—survival time—chosen with planned Bayesian analysis
- Sample size of 400 justified via reverse philosophy notion[22]

because a non-significant result will be inevitable if such an effect size is unfeasible. However, thought should be given to the type of treatment and the population for the trial to maximise the size of the feasible treatment effect.

If the rare cancer is associated with a targeted driver mutation or the rarity of the cancer is a result of stratification of a more common cancer by a predictive biomarker, expectation of a larger treatment effect size, thereby permitting reduction in sample size, is rational. This principle is key to enrichment of trial designs.[23] The disadvantage of reducing the pool of patients has to be weighed against the reduction in sample size that results from a design with an expected large treatment effect.

Compromise on phase 3 trial design convention

Relaxation of type I and type II error rates to allow them to be larger than the conventional values is a recognised compromise of the statistical design to reduce sample size—eg, setting the significance level to 10% rather than 5% (ie, accepting a one in ten chance that the trial has wrongly yielded a positive result, thus erroneously changing practice), reducing power to 80% rather than 90% (ie, accepting a one in five chance of wrongly missing a true positive result and thus failing to change practice appropriately). Patients, clinicians, funders, and regulatory authorities then have to decide whether such evidence is sufficiently robust. Furthermore, the design can be based on one-sided hypothesis testing in preference to the conventional two-sided testing to reduce the sample size further. This approach was implemented in the International Rare Cancers Initiative trial design for salivary gland cancer[22] (annual incidence 2·5–3·0 per 100 000 people; IRCI 007, NCT01969578), in which sample size was based on a one-sided significance level of 10% and power of 80%. Even a small randomised trial could provide valuable unbiased evidence to inform clinical practice for rare cancers, despite being less robust than conventional designs, especially because it could be included in meta-analyses.[24]

Simple and small or complex with maximum utility?

The plan to recruit a rare cancer population to a pivotal clinical trial might be seen as a unique opportunity to assess interventions at several points in a treatment pathway or compare many possible interventions in parallel to maximise the information gained—eg, a multiarm multistage design.[25] Although such designs might be efficient compared with several simple two-group randomised controlled trials, the required sample size for more complex designs might be prohibitively large in a rare cancer population.

In comparison to a simple randomised controlled trial with two parallel groups (ie, experimental *vs* control), designs with more than two treatment groups necessitate an increased total sample size not only because of the additional groups that patients need to be recruited to but also to adjust for type I errors for multiple comparisons in the conventional frequentist setting. If the design also incorporates several interim analyses rather than a single, planned final analysis, then further adjustment of the type I errors will be necessary, thereby increasing the required sample size further. The advantages of assessing several interventions in one trial need to be weighed against the disadvantages of spreading a small population even more

thinly across randomised treatment options. One approach is to set a simple single randomised comparison within a wider registration study: prospective data collection can be valuable even if a hypothesis is not tested, and establishment of a collaborative framework for the pivotal trial could also catalyse development of further independent trials.

External rather than internal controls

An extreme form of design simplification is to do a single-group trial rather than a randomised controlled trial and use controls external to the trial protocol as a comparison. This design is widely used for phase 2 trials to justify proceeding to a subsequent phase 3 trial with conventional methods, such as A'Hern's single stage design or Simon's two stage design. In particular circumstances when the external control data are robust, use of such designs with more stringent statistical error rates might allow a single-group trial to change clinical practice. For example, Cancer Research UK approved and funded a single-group phase 3 trial in non-seminomatous germ-cell tumour of the testis (annual incidence for all testicular cancer is seven per 100 000) because the historical control data for treatment benefit were so well established and robustly close to 100% (the 111 trial, CRUK/09/011, NCT01726374, figure 3, panel 2). A major advantage of single-group trials is that fewer participants are needed than for a randomised controlled trial, but the disadvantage is that the resulting evidence is perceived as being of lower quality for informing future practice. The European Medicines Agency and FDA acknowledge that this approach might be acceptable under exceptional circumstances.[16,17] This level of evidence can be accepted as sufficient in the rare cancer setting on the basis of a review[19] of FDA approvals for 68 different rare cancer indications; 66 of the 99 trials submitted as supporting evidence were single-group trials.

Selection of outcome measurements

Selection of outcome measures for cancer trials should always be considered carefully,[26] and especially for rare cancers, for which it directly affects the sample size (table 1).[27] For people with a life-threatening and progressive disease such as cancer, overall survival—or sometimes progression-free survival—is conventionally thought to be the most meaningful outcome in trials to change clinical practice, with recurrence-free survival important for early-stage disease. If the effect of treatment on this outcome is expected to be smaller than can be detected through a feasible sample size, use of an alternative outcome measure might be preferable—eg, one that is less clinically relevant but still important, which would necessitate a smaller sample size, either because it results in a bigger treatment effect or in a time-to-event context it has a greater event rate.

Validated surrogate outcome measures that relate to survival time would be the ideal choice but are not easy to find.[28] Surrogates such as response are legitimate in phase 2 trials and have been sufficient to change clinical practice in rare cancers: 69% of FDA approvals for rare cancer indications are based on the primary efficacy endpoint of objective response.[19] Use of continuous measures—eg, change in tumour size—can result in

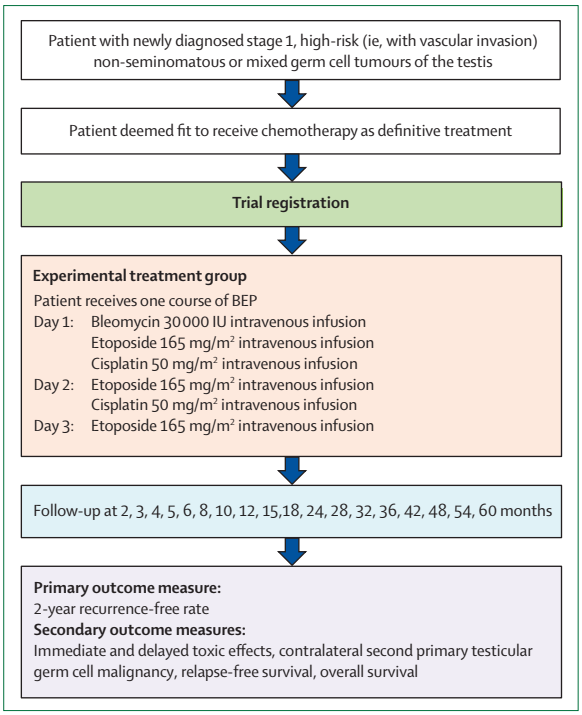

Figure 3: **An uncontrolled phase 3 rare cancer trial: the 111 trial**
Trial registration: CRUK/09/011; NCT01726374. BEP=bleomycin, etoposide, and cisplatin. IU=international units.

Panel 2: **Statistical design for phase 3 trial for a rare cancer: the 111 trial**

- Standard treatment: two courses of BEP (bleomycin, etoposide, cisplatin)
- Historical data show 2-year recurrence free rate of 98% on standard treatment (MRC TE05 trial)
- Historical control data are robust and well established: thus, randomisation of patients to a control group was deemed wasteful
- Trial had single-group single-stage phase 2 trial design by A'Hern
- Benchmark comparison from 2-year recurrence-free rate of 95% deemed to be the minimum acceptable level for the experimental treatment
- Statistical error rates set as α=5% and β=20%
- Trial is designed to have 80% chance of demonstrating that the 95% one-sided confidence interval for 2-year recurrence-free rate excludes 95% if the true rate is 98%
- Trial requires recruitment of 236 patients and maximum of seven recurrences noted within 2 years to exclude the 2-year recurrence-free rate of 95%

greater statistical power than an associated dichotomous variable such as response and thus reduced sample size is needed.[29]

Trial investigators should aim to maximise the information collected for each patient, including measuring various outcome measures at relevant timepoints during treatment and follow-up—ideally with an associated prespecified hierarchy of importance for decision making. In the rare cancer setting, particularly for extremely rare malignancies, collection and presentation of data for all potentially relevant outcomes might be preferable to prespecification of a primary outcome on which decisions are made. If collectively the data look compelling, then that could be sufficient to change clinical practice.[16]

Reduce noise

A key principle when doing a clinical trial in rare cancer is maximisation of the usefulness of the data gathered in the trial, to enable a risk–benefit assessment.[16] When the sample size is restricted, the quality of the data becomes particularly important to detect effects above noise. Trial conduct should ensure that missing data and protocol violations are kept to a minimum and that patients are completely followed up.

One way to make the most of the data gathered is to adjust for baseline characteristics when comparing treatment groups in terms of outcome measures by using statistical modelling. Such an approach controls the variability in the outcome measure and thereby increases the statistical power but might not be feasible in the rare cancer setting if the study size is small.

Adaptive designs

An alternative to a fixed sample size in a conventional trial design is an adaptive trial design, in which use of information gained from interim analyses to make adaptations while in progress is prospectively planned. A wide range of adaptations are possible.[30] Adaptive design aims to minimise the number of patients needed to maintain statistical validity, thereby improving the efficiency of trials. Thus, it is an approach mentioned in most published methodological reviews in rare diseases, although not necessarily applicable in the context of trials aimed to change clinical practice in rare cancers (table 2). Widely cited examples of adaptive designs in cancer trials include the BATTLE trial[31] (adaptive randomisation[32]) and STAMPEDE[33] (multiarm multistage design[25]), but the benefits in the rare cancer setting are debatable. The sample size for an adaptive design is generally unknown in advance because it is dependent on the results of the interim analyses, but in theory adaptive designs are more efficient in the long run because the expected sample size for the continuously assessed adaptive design would be smaller than for the corresponding fixed sample size. However, there is no guarantee in the single trial being planned that the sample size will be smaller than a design with a fixed sample size, and planning a trial with an unknown sample size in a rare cancer setting might be too big a risk to take. Additionally, undertaking a trial in a rare cancer could be thought of as a precious opportunity to obtain valuable information, and thus running a trial with a feasible fixed sample size might be preferable to an adaptive design, to maximise that opportunity.

Alternative designs to change clinical practice

Bayesian methods

Bayesian methods have been promoted for clinical trials in general,[34,35] but are still widely thought of as unconventional. They are advocated as an option for trials in rare diseases in reviews[4–10,12,13] and by the European Medicines Agency[16] (table 1), and are being used in three International Rare Cancers Initiative trials.[22] Bayesian methods are an alternative to conventional frequentist statistical analysis and were proposed by one of the earliest papers to discuss designing trials for rare diseases.[36] The Bayesian approach was also promoted at that time more generally in relation to small clinical trials.[37]

If one starts from the premise before a trial that substantial uncertainty exists in relation to the unknown size of a treatment effect, then the philosophy underpinning the Bayesian approach is to abandon attempts to provide a definitive answer through hypothesis testing and use the unbiased evidence gathered in a well designed clinical trial (even in only a small sample) as a way of reducing that uncertainty. This approach could help clinicians to make treatment decisions, particularly when it provides a more intuitive probabilistic approach to reporting results.

When providing an estimation of the size of a treatment effect, the Bayesian approach allows external and subjective information, expressed as a prior probability distribution, to be combined with trial evidence to give a posterior probability distribution. This approach ensures that the trial is reducing uncertainty about the treatment effect from a level of knowledge that already exists. Further to the proposal by Lilford and colleagues,[36] a strategy was developed for designing trials to assess interventions in rare cancers, specifically in terms of survival time as an outcome measure.[38,39] The strategy contained methods for the creation of a prior distribution from existing evidence. It suggests searching published work for all evidence relating to a proposed trial, even including studies with only tentative similarities in terms of type of cancer, treatment, and endpoints, and all levels of evidence—from randomised controlled trials to single case study reports. This evidence can then be combined into a prior distribution for the treatment effect, with weights allocated in relation to pertinence, validity, and precision. In principle, this idea is sensible, but in practice such an approach is problematic because, in the rare cancer setting, previous evidence will generally be low level and therefore developing a plausible prior can be problematic.[40]

One could argue that clinical decision makers will implicitly use a Bayesian approach because they do not interpret trial evidence in a vacuum, but rather take into account previous knowledge and preconceived notions about the effectiveness of the treatment.

If the development or acceptance of prior probability distributions is problematic then the Bayesian approach can be implemented with non-informative prior distributions (ie, when all treatment effects are roughly equally likely), which allows a trial's results to be expressed in terms of direct probabilities of the treatment effect being a certain size—one of the great advantages of the method. For example, the sort of result that one would be able to conclude in terms of a survival outcome is that, in view of previous evidence (which might not exist) and the trial data, there is a 70% chance that the treatment truly reduces the hazard of death by at least 10% (ie, HR<0·9). This type of reporting could be used practically by clinicians in discussion with patients and enable evidence-based treatment decisions, while conversely, a non-significant result from hypothesis testing would simply be thought of as inconclusive or, at worst, incorrectly interpreted as evidence of no treatment effect.

The trials of the International Rare Cancers Initiative show how Bayesian approaches can be used, with designs based entirely around a Bayesian approach or a planned Bayesian analysis supplementing the main frequentist design.[22] In the randomised controlled trial in which androgen deprivation therapy was compared with chemotherapy in recurrent or metastatic androgen-receptor-expressing salivary gland cancer (annual incidence of 2·5–3·0 per 100 000; IRCI 007, NCT01969578), Bayesian methods are used as secondary sensitivity analysis. The design is based on a conventional frequentist analysis of progression-free survival time, with relaxed statistical error rates but the investigators intend to assess the robustness of the conclusions with a Bayesian analysis with various prior distributions for the progression-free survival HR.

The randomised controlled trial of adjuvant therapy in stage I–III small bowel adenocarcinoma (annual incidence 2·2–5·7 per 100 000; BALLAD, IRCI 002, NCT02502370) mixes frequentist and Bayesian inferential frameworks. The planned sample size is based on a conventional frequentist analysis of disease-free survival with relaxed statistical error rates. If the trial produces a significant result, then the final estimates of the treatment effect will be generated with a Bayesian analysis to combine the trial data with clinicians' consensus prior beliefs based on interpretation of relevant external evidence. The randomised controlled trial of locally advanced squamous carcinoma of the penis (annual incidence 1·5 per 100 000; InPACT, CRUK/13/005, IRCI 004, NCT02305654) is based on a wholly Bayesian approach with a reverse philosophy to sample size (figure 2). Unlike conventional designs, the design starts with the number of patients that can be feasibly recruited within a sensible timeframe and then assesses whether that amount of data would have sufficient value in terms of reducing uncertainty about a minimum treatment effect size to justify a trial. Such a philosophy has been previously introduced in the frequentist context.[41]

Observational studies: capturing retrospective data

Observational studies based on large, multicentre databases have the potential to describe characteristics, interventions, and outcomes for patients with rare cancers caught within their scope. They include administrative databases used for surveillance and epidemiology and those used to manage reimbursement. Some causal inference can be interpreted, although it will always be subject to several sources of bias and confounding. Providing that the data have been captured, several variables that can affect treatment outcomes can be analysed. For example, the National Cancer Institute's Surveillance, Epidemiology and End Results programme (SEER) collects data from about 26% of the US population, including demographics, primary tumour, and first treatment. It is linked to the Medicare database, which captures initial diagnosis and medical care for patients older than 65 years.

The SEER registry has been extensively mined to inform treatment guidelines for rare cancers. For example, studies of SEER[42,43] have shown increased risks of Merkel cell carcinoma, a rare, aggressive malignancy of older patients (median age 78 years), in patients with autoimmune conditions. Merkel cell carcinoma is locally invasive and has high metastatic potential. Sentinel lymph node biopsy is widely used to guide regional nodal management. Retrospective analysis of 5-year Merkel-cell-carcinoma-specific survival in 1193 patients showed not only that patients with negative sentinel lymph node biopsy results had much better outcomes than those with positive results but also that patients undergoing sentinel lymph node biopsies had better outcomes than those who did not.[44] Such observational data are used to support the incorporation of sentinel lymph node biopsy into guidelines[45] but cannot account for confounding factors such as comorbidities within an elderly population that might affect the decision to undergo an operative procedure.

The next generation of large observational databases is likely to arise from the evolution of electronic patient records, and should overcome the diversity of platforms between medical institutions and adapt existing designs that enable mining of large volume data from the commercial and financial sectors to real-world cancer data. CancerLinq is a prototype developed by the American Society of Clinical Oncology to extract data continuously from the electronic records of patients with cancer that is structured within a single searchable platform. The use of natural language processing enables extraction from medical notes rather than limiting

For more on **CancerLinq** see http://cancerlinq.org

collection to prespecified fields. The purpose is to use continuously updated data to generate and modify decision-making algorithms, providing point-of-care data to support clinical practice and patients' confidence as well as a data source for researchers and guideline development.[46] In the UK, care.data is a programme in development to collect rounded data from National Health Service patients. It aims to be the basis for research into the prevention and management of disease (ie, not limited to cancer) and to support the efficient and fair planning of services.[47] Important concerns being addressed include patients' right to opt-out, and ensuring that the scheme collects only anonymised coded medical data (and not, for example, notes of conversations with primary care doctors) and safeguards against purely commercial use of data, such as setting insurance premiums. The risks and benefits must be communicated effectively with all sectors of society.

Observational datasets are retrospective, non-randomised, and not governed by protocols, which means that analysis of outcomes in relation to intervention is compromised. Established methods[48] address selection bias by adjusting analysis for known covariates. These methods include matching treated with untreated patients by one or more covariate, quantifying the conditional probability that a patient would receive the intervention on the basis of many covariates (ie, propensity scoring[49]), and building models that adjust for covariates and propensity scores. Several methods have been developed to detect, quantify, and—possibly—adjust for hidden bias arising from unknown covariates.[50]

Databases of real-world data will allow capture and grouping of more detailed data for management of rare cancer populations than do currently available databases. They greatly extend systematic clinical observation from the small and potentially unrepresentative population who participate in clinical trials. If trials have not been done because of the paucity of patients, real-world data are likely to become a valuable resource in the development of standards of care. Use of real-world data is being advanced as a solution to the increasing scale of data for targets and potential treatments for cancers characterised by molecular signatures. However, prospective trials remain essential for rare cancer populations because of the inherent bias in the interpretation of observational data, and thus the complementary role of databases of real-world data alongside data generated in clinical trials needs to be established.[51]

Singletons, baskets, and umbrellas

The n-of-1 trial[52] is widely described as a research method in rare diseases (table 1), but its relevance to and application in rare cancer populations have hardly been explored. When we searched for trials on PubMed with the terms "n-of-1" and "trials", we identified 83 reports, which were overwhelmingly of palliative interventions in the setting of non-malignant chronic diseases. Six of the studies were of supportive care interventions in cancer treatment. ClinicalTrials.gov lists 20 trials under "n-of-1" for similar indications (ie, non-malignant chronic diseases and supportive care in cancer treatment), except for one (NCT02142036) of targeted therapy in the advanced cancer setting.

The n-of-1 trial differs from a case report in that it is typically a prospective sequence of interventions randomly assigned in one patient (rather than patients being randomly allocated to a treatment group as in a conventional comparative study). Several interventions and placebo can be used and switches between alternatives can be made. The simple conclusion of an n-of-1 trial is the selection of the best intervention for that individual. Thus, n-of-1 trials primarily serve as a formalised mechanism of individualised decision making in which the participant is both trial participant and direct beneficiary of the analysis. However, wider inferences can be drawn when data from serial related n-of-1 trials of similar patients are aggregated retrospectively—or, ideally, prospectively, as a planned experiment providing more convincing evidence.[53] Serial n-of-1 trials in patients separated in time permits adaptation of the protocol based on previous response data in patients with tumours that have similar molecular aberrations.

What are the roles and characteristics of formal prospective n-of-1 trial designs for interventions to treat malignancy (as opposed to the control of chronic symptoms) for rare cancer populations? The standard design would be meaningful and ethical only if the malignancy were following a chronic and indolent course, if it seemed likely that the malignancy were reverting to its previous natural history between treatment blocks, and if there were meaningful short-term outcome measures of response. Because intervention might change a cancer's molecular characteristics, it is unlikely to be possible or meaningful to randomly assign a sequence of treatment versus placebo for a disease that is relentlessly progressive and symptomatic. A more practical version of the n-of-1 trial for cancer would be to compare response and the kinetics of tumour growth for a single block of experimental treatment as per protocol against the predicted outcomes with standard treatment—ie, an uncontrolled n-of-1 trial. Such predictions might be based on data taken from the same patient before recruitment or historical data for patients given standard therapy, although few such data might be available. Such a prospective, systematic, and iterative framework comprising several unique patient-specific treatment trials could perhaps affect clinical practice for rare cancers.

Patients with diverse malignancies can be recruited on the basis of sharing a molecular classification, to basket trials, in which one or more personalised interventions is tested systematically within a single protocol, often

with a common molecular screening platform and typically in a multicentre cooperative group (figure 4).[23,54] No comparison is intended between the different types of cancer, and treatment success might vary between cancers because of the different biological contexts of the common molecular target. Assessment of the intervention can be made within each tumour type and across the whole basket. This design can be an efficient way to do trials for rare cancer populations, and overcomes the hurdle of investment that would blight a single-disease study with poor accrual.

Basket trials make sense in settings in which the treatment effect is expected to be large on the basis of biological selection of patients by expression of a putative drug target supported by laboratory evidence of relevance to oncogenesis and inhibition by the drug. Allocation can be done via a molecular tumour board, which can be used to assess the relative priority of different targets and drugs using previous preclinical and clinical data. This set-up is related to the principles of n-of-1 trials: the basket trial offers a framework for individualised decision making in which the patient is both trial participant and direct beneficiary of the analysis. The National Cancer Institute's Molecular Analysis for Therapy Choice is a phase 2 trial (NCT02465060) with an aim of recruiting more than 3000 patients with diverse refractory cancers; at least 1000 patients will be allocated into 20–25 small molecularly defined cohorts (n=~35) and given a relevant drug with proven efficacy in some setting. It is designed such that at least a quarter of enrolled patients will have malignant disease other than the most common small cell lung, prostate, breast, or colon cancers. Clinical response and 6-month progression-free survival are the outcomes measured with interpretation based on nominal low frequencies for historical controls.

Basket trials are thought of as discovery trials aimed at efficiently screening drugs for efficacy across several tumour settings. They help to reduce the uncertainty surrounding the relevance of a molecular target in different types of tumour. Although basket trials are hypothesis generating for common cancers, for rare cancer populations they could be an efficient mechanism for generating data that can directly affect clinical practice. In our view, for the trial to potentially affect future treatment of biologically similar patients, consistency in several outcomes should be reported, including reduction of tumour burden, pharmacodynamic evidence of target inhibition, and improvement in patient-reported outcomes, such as clear improvement in levels of activity.

Whereas basket trials work for rare cancer populations by grouping them with others according to shared molecular targets, umbrella trials split tissue-defined cancer populations into molecularly defined subgroups, some of which might thus be rare (figure 4).[23,54] Similar to basket trials, umbrella trials are multi-institution projects, typically with a common molecular screening

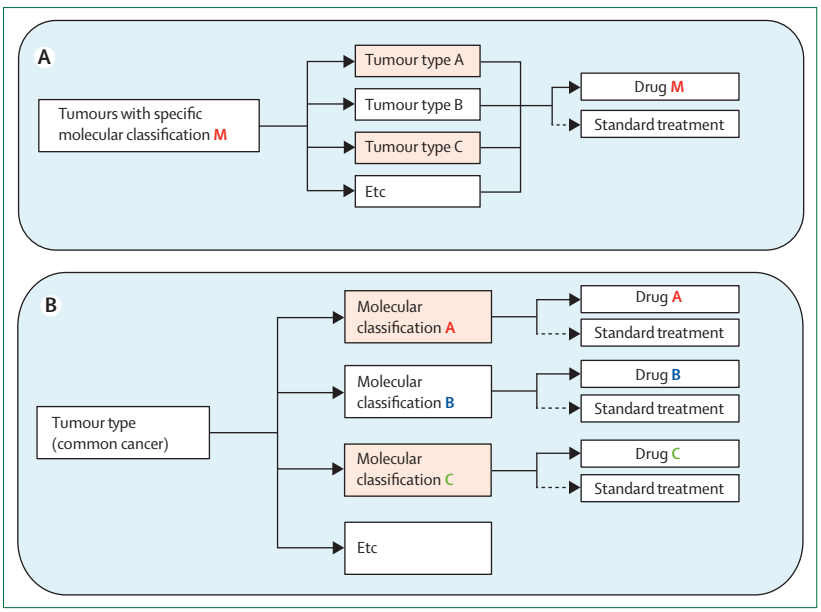

Figure 4: **Use of basket trials to assess an experimental drug in tumours with a specific molecular classification (A); and umbrella trials to assess several experimental drugs in a range of molecular classifications within a tumour type (B)**

Dotted lines suggest possible randomised trial design rather than uncontrolled design. The shaded boxes represent rare populations, and show the potential application of these designs to rare cancers. Drugs M, A, B, and C are experimental drugs targeted to a specific molecular classification (M, A, B, and C, respectively), which have come about because either the drug has been developed for that specific molecular target or it has been shown to be effective in that particular group of patients.

platform. The structure creates economy of scale, which makes offering intervention to a rare molecularly defined group feasible. The logistics of such complex trials—eg, development of a common molecular screening platform and work across pharmaceutical companies to supply a range of targeted drugs—are challenging. Umbrella trials can be viewed as a group of enrichment trials with new groups starting or being dropped with time. They are typically later phase than basket trials, and each enrichment trial might be single group or include randomisation of the experimental intervention and standard of care. Umbrella trials can be designed as phase 2 to establish whether to proceed to phase 3 or designed to determine standard of care in the stratified population. In some settings, even a single-group trial can change practice if the rationale for selection of patients is strong and targeted intervention results in a sizeable treatment effect.

Conclusion

Methods advocated for clinical trials in rare diseases are not necessarily applicable in rare cancers, and the choice of method used is related to where the cancer lies on a sliding scale of rarity. The approaches that we discussed will enable evidence-based clinical practice for rare cancers in the future, particularly in the rapidly developing era of molecularly defined cancers. Methods outside the conventional approach might not be ideal, but pragmatic

Search strategy and selection criteria

We searched MEDLINE with index terms "rare diseases", "clinical trials as topic", "research design", and "humans" on May 15, 2015. We did not limit our search by date or language of publication. Our search returned 133 unique references, which were screened by LB for relevance. Articles were deemed relevant if they described and discussed methods for late phase clinical trials that would be applicable for rare diseases. We also identified additional articles through searches of our files. Priority for inclusion in the final reference list was given to easily accessible references, more recent publications, and articles that were cancer-specific and substantial articles rather than commentaries or editorials.

solutions have to be sought to enable some level of evidence-based health care in this setting. Exploration of methods for the design and analysis of clinical studies of rare diseases continues to be a research priority, as evidenced by three projects funded by the European Union within the 7th Framework Programme (ASTERIX: FP7 HEALTH 2013–603160; IDEAL: FP7 HEALTH 2013–602552; INSPIRE: FP7 HEALTH 2013–602144). Such methodological developments could further improve assessments of new treatments in rare cancers and ultimately improve patient care.

Contributors

LB and NS contributed equally to the drafting of the manuscript, tables, and figures. KM did the search of published work and reviewed and commented on drafts of the manuscript.

Declaration of interests

We declare no competing interests.

Acknowledgments

This research was supported by Cancer Research UK (through funding salaries for LB and KM), the European Network for Cancer Research in Children and Adolescents (FP7-HEALTH-F2–2011 contract number 261474), and the Birmingham Experimental Cancer Medicine Centre. Preliminary work for this paper was done as part of the Medical Research Council Midland Hub for Trials Methodology Research (grant number G0800808). We thank Steve Nicholson (international study chair for the InPACT trial), Michael Cullen (chief investigator of the 111 trial), and Emma Hall (as lead statistician and coinvestigator of both trials) for their permission to use their trial flow diagrams.

References

1 Gatta G, van der Zwan JM, Casali PG, et al, and the RARECARE working group. Rare cancers are not so rare: the rare cancer burden in Europe. *Eur J Cancer* 2011; **47:** 2493–511.

2 Garraway LA. Genomics-driven oncology: framework for an emerging paradigm. *J Clin Oncol* 2013; **31:** 1806–14.

3 Lawrence MS, Stojanov P, Polak P, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* 2013; **499:** 214–18.

4 Gagne JJ, Thompson L, O'Keefe K, Kesselheim AS. Innovative research methods for studying treatments for rare diseases: methodological review. *BMJ* 2014; **349:** g6802.

5 Cornu C, Kassai B, Fisch R, et al, and the CRESim & Epi-CRESim Project Groups. Experimental designs for small randomised clinical trials: an algorithm for choice. *Orphanet J Rare Dis* 2013; **8:** 48.

6 Gupta S, Faughnan ME, Tomlinson GA, Bayoumi AM. A framework for applying unfamiliar trial designs in studies of rare diseases. *J Clin Epidemiol* 2011; **64:** 1085–94.

7 Kianifard F, Islam MZ. A guide to the design and analysis of small clinical studies. *Pharm Stat* 2011; **10:** 363–68.

8 Griggs RC, Batshaw M, Dunkle M, et al, and the Rare Diseases Clinical Research Network. Clinical research for rare disease: opportunities, challenges, and solutions. *Mol Genet Metab* 2009; **96:** 20–26.

9 Buckley BM. Clinical trials of orphan medicines. *Lancet* 2008; **371:** 2051–55.

10 Behera M, Kumar A, Soares HP, Sokol L, Djulbegovic B. Evidence-based medicine for rare diseases: implications for data interpretation and clinical trial design. *Cancer Control* 2007; **14:** 160–66.

11 Lagakos SW. Clinical trials and rare diseases. *N Engl J Med* 2003; **348:** 2455–56.

12 Augustine EF, Adams HR, Mink JW. Clinical trials in rare disease: challenges and opportunities. *J Child Neurol* 2013; **28:** 1142–50.

13 Tudur Smith C, Williamson PR, Beresford MW. Methodology of clinical trials for rare diseases. *Best Pract Res Clin Rheumatol* 2014; **28:** 247–62.

14 Van der Lee JH, Wesseling J, Tanck MWT, Offringa M. Efficient ways exist to obtain the optimal sample size in clinical trials in rare diseases. *J Clin Epidemiol* 2008; **61:** 324–30.

15 Wilcken B. Rare diseases and the assessment of intervention: what sorts of clinical trials can we use? *J Inherit Metab Dis* 2001; **24:** 291–98.

16 European Medicines Agency, Committee for Medicinal Products for Human Use. Guideline on clinical trials in small populations. http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003615.pdf (accessed Jan 14, 2015).

17 US Department of Health and Human Services, Food and Drug Administration. Rare diseases: common issues in drug development guidance for industry; draft guidance August, 2015. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM458485.pdf (accessed Jan 14, 2015).

18 Casali PG. Do rare cancers deserve specific strategies for cancer research? *Lancet Oncol* 2010; **11:** 506–07.

19 Gaddipati H, Liu K, Pariser A, Pazdur R. Rare cancer trial design: lessons from FDA approvals. *Clin Cancer Res* 2012; **18:** 5172–78.

20 Casali PG, Bruzzi P, Bogaerts J, Blay JY, and the Rare Cancers Europe (RCE) Consensus Panel. Rare Cancers Europe (RCE) methodological recommendations for clinical studies in rare cancers: a European consensus position paper. *Ann Oncol* 2015; **26:** 300–06.

21 Keat N, Law K, Seymour M, et al. International Rare Cancers Initiative. *Lancet Oncol* 2013; **14:** 109–10.

22 Bogaerts J, Sydes MR, Keat N, et al. Clinical trial designs for rare diseases: studies developed and discussed by the International Rare Cancers Initiative. *Eur J Cancer* 2015; **51:** 271–81.

23 Mandrekar SJ, Dahlberg SE, Simon R. Improving clinical trial efficiency: thinking outside the box. *Am Soc Clin Oncol Educ Book* 2015; **35:** e141–47.

24 Halpern SD, Karlawish JH, Berlin JA. The continuing unethical conduct of underpowered clinical trials. *JAMA* 2002; **288:** 358–62.

25 Royston P, Parmar MK, Qian W. Novel designs for multi-arm clinical trials with survival outcomes with an application in ovarian cancer. *Stat Med* 2003; **22:** 2239–56.

26 Wilson MK, Collyar D, Chingos DT, et al. Outcomes and endpoints in cancer trials: bridging the divide. *Lancet Oncol* 2015; **16:** e43–52.

27 Sleijfer S, Wagner AJ. The challenge of choosing appropriate end points in single-arm phase II studies of rare diseases. *J Clin Oncol* 2012; **30:** 896–98.

28 Schatzkin A, Gail M. The promise and peril of surrogate end points in cancer research. *Nat Rev Cancer* 2002; **2:** 19–27.

29 Wason JM, Mander AP, Eisen TG. Reducing sample sizes in two-stage phase II cancer trials by using continuous tumour shrinkage end-points. *Eur J Cancer* 2011; **47:** 983–89.

30 Chow SC, Chang M. Adaptive design methods in clinical trials—a review. *Orphanet J Rare Dis* 2008; **3:** 11.

31 Kim ES, Herbst RS, Wistuba II, et al. The BATTLE trial: personalizing therapy for lung cancer. *Cancer Discov* 2011; **1:** 44–53.

32 Korn EL, Freidlin B. Outcome—adaptive randomization: is it useful? *J Clin Oncol* 2011; **29:** 771–76.

33 Sydes MR, Parmar MK, Mason MD, et al. Flexible trial design in practice—stopping arms for lack-of-benefit and adding research arms mid-trial in STAMPEDE: a multi-arm multi-stage randomized controlled trial. *Trials* 2012; **13:** 168.

34 Berry DA. Bayesian clinical trials. *Nat Rev Drug Discov* 2006; **5:** 27–36.

35 Spiegelhalter DJ, Abrams KR, Myles JP. Bayesian approaches to clinical trials and health evaluation: Chichester, John Wiley & Sons, 2004.

36 Lilford RJ, Thornton JG, Braunholtz D. Clinical trials and rare diseases: a way out of a conundrum. *BMJ* 1995; **311:** 1621–25.

37 Matthews JN. Small clinical trials: are they all bad? *Stat Med* 1995; **14:** 115–26.

38 Tan SB, Dear KB, Bruzzi P, Machin D. Strategy for randomised clinical trials in rare cancers. *BMJ* 2003; **327:** 47–49.

39 Tan SB, Wee J, Wong HB, Machin D. Can external and subjective information ever be used to reduce the size of randomised controlled trials? *Contemp Clin Trials* 2008; **29:** 211–19.

40 Billingham L, Malottki K, Steven N. Small sample sizes in clinical trials: a statistician's perspective. *Clin Invest* 2012; **2:** 655–57.

41 Parker RA, Berman NG. Sample size: more than calculations. *Am Stat* 2003; **57:** 166–70.

42 Lanoy E, Engels EA. Skin cancers associated with autoimmune conditions among elderly adults. *Br J Cancer* 2010; **103:** 112–14.

43 Lanoy E, Costagliola D, Engels EA. Skin cancers associated with HIV infection and solid organ transplant among elderly adults. *Int J Cancer* 2009.

44 Kachare SD, Wong JH, Vohra NA, Zervos EE, Fitzgerald TL. Sentinel lymph node biopsy is associated with improved survival in Merkel cell carcinoma. *Ann Surg Oncol* 2014; **21:** 1624–30.

45 Bichakjian CK, Olencki T, Alam M, et al. Merkel cell carcinoma, version 1.2014. *J Natl Compr Canc Netw* 2014; **12:** 410–24

46 Schilsky RL, Michels DL, Kearbey AH, Yu PP, Hudis CA. Building a rapid learning health care system for oncology: the regulatory framework of CancerLinQ. *J Clin Oncol* 2014; **32:** 2373–79.

47 NHS England. The care.data programme—collecting information for the health of the nation. http://www.england.nhs.uk/ourwork/tsd/care-data/ (accessed June 4, 2015).

48 Deeks JJ, Dinnes J, D'Amico R, et al, and the International Stroke Trial Collaborative Group, and the European Carotid Surgery Trial Collaborative Group. Evaluating non-randomised intervention studies. *Health Technol Assess* 2003; **7:** iii–x, 1–173.

49 D'Agostino RB Jr. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Stat Med* 1998; **17:** 2265–81.

50 Alemayehu D, Alvir JM, Jones B, Willke RJ. Statistical issues with the analysis of nonrandomized studies in comparative effectiveness research. *J Manag Care Pharm* 2011; **17** (suppl A)**:** S22–26.

51 Black N. Why we need observational studies to evaluate the effectiveness of health care. *BMJ* 1996; **312:** 1215–18.

52 Senn S. n-of-1 trials. In: Statistical issues in drug development, 2nd edn. Chichester: John Wiley & Sons Ltd, 2008: 287–94.

53 Zucker DR, Ruthazer R, Schmid CH. Individual (n-of-1) trials can be combined to give population comparative treatment effect estimates: methodologic considerations. *J Clin Epidemiol* 2010; **63:** 1312–23.

54 Menis J, Hasan B, Besse B. New clinical research strategies in thoracic oncology: clinical trial design, adaptive, basket and umbrella trials, new end-points and new evaluations of response. *Eur Respir Rev* 2014; **23:** 367–78.