Title

Progression-free survival, disease-free survival and other composite end points in oncology: improved reporting is needed

Journal

Nature Reviews Clinical Oncology, 20(12)

ISSN

1759-4774 1759-4782

Authors

Walia, Anushka
Tuia, Jordan
Prasad, Vinay

Publication Date

2023-10-12

DOI

10.1038/s41571-023-00823-5

Peer reviewed

nature reviews clinical oncology

https://doi.org/10.1038/s41571-023-00823-5

Perspective

Check for updates

Progression-free survival, disease-free survival and other composite end points in oncology: improved reporting is needed

Anushka Walia [1] ✉, Jordan Tuia [2] & Vinay Prasad [2]

Abstract

Composite outcome measures such as progression-free survival and disease-free survival are increasingly used as surrogate end points in oncology research, frequently serving as the primary end point of pivotal trials that form the basis for FDA and EMA approvals. Such outcome measures combine two or more distinct events (for example, tumour (re)growth, new lesions and/or death) into a single, time-to-event end point. The use of a composite end point can increase the statistical power of a clinical trial and decrease the follow-up period required to demonstrate efficacy, thus lowering costs; however, these end points have a number of limitations. Composite outcomes are often vaguely defined, with definitions that vary greatly between studies, complicating comparisons of results across trials. Altering the makeup of events included in a composite outcome can alter study conclusions, including whether treatment effects are statistically significant. Moreover, the events included in a composite outcome often vary in clinical significance, reflect distinct biological pathways and/or are affected differently by treatment. Therefore, knowing the precise breakdown of the component events is essential to accurately interpret trial results and gauge the true benefit of an intervention. In oncology clinical trials, however, such information is rarely provided. In this Perspective, we emphasize this deficiency through a review of 50 studies with progression-free survival as an outcome published in five top oncology journals, discuss the advantages and challenges of using composite end points, and highlight the need for transparent reporting of the component events.

Sections

A full list of affiliations appears at the end of the paper. ✉e-mail: anushka.walia@ucsf.edu

Introduction

Progression-free survival (PFS), disease-free survival (DFS), relapse-free survival (RFS) and event-free survival (EFS) are among the most common end points used in oncology clinical trials and have gained popularity in recent decades[1]. All of these outcome measures are time-to-event end points, meaning that they are defined as the time from randomization or registration of the patient to an event of interest, such as death or disease progression[2]. Additionally, they are all composite end points — groupings of two or more distinct clinical events that are pooled into one aggregate outcome[3]. Composite end points have a long history in biomedical research and are also frequently used in cardiology and nephrology clinical trials (for example, major adverse cardiac events (MACE), as a composite of non-fatal stroke, non-fatal myocardial infarction, cardiovascular disease-related death and, in some cases, other events such as death from any cause and hospitalization, in cardiology). In these disciplines, reporting of not only the aggregate result, but also each component outcome separately is customary. Notably, this distinction of individual component events seldom occurs in oncology, with clinical trial reports typically providing data on only the composite end point as a whole. In this Perspective, we summarize the history relating to the adoption of composite end points in oncology and related fields, define composite outcomes that are used across different cancer types and treatment settings, describe the specific strengths and risks of composite end points, and explain why clinical research should routinely report composite events both in aggregate and separately. Such transparency would enable a better understanding of cancer biology and clinical trial results, and ultimately improve patient care.

Overview of composite end points

Composite end points are widely used in clinical research as outcomes delineating the time to first occurrence of any one of the multiple component events in each patient included in the study cohort. Therefore, by increasing the number of events captured in a single outcome, composite end points can increase the statistical power of a study, thus reducing the sample size, follow-up time and financial outlay required to demonstrate a clinical effect. Despite these advantages, the use of composite end points is controversial owing to inconsistent definitions, heterogeneity of component events, differing levels of clinical importance among the various components, and potentially misleading interpretations.

Composite end points are used in almost all fields of biomedicine. In cardiology, MACE is the most recognized composite outcome. In oncology, composite end points such as PFS and DFS are frequently used as surrogates for 'true', clinically meaningful end points including overall survival (OS) and/or health-related quality of life (HRQOL). The use of such composite end points is increasing in correlation with the number of anticancer agents approved. In an analysis of randomized controlled trials (RCTs) focused on breast cancer, colorectal cancer and non-small-cell lung cancer (NSCLC), 125 (42%) of 298 RCTs published in seven major journals between 2010 and 2020 used PFS as a primary end point, compared with 25 (18%) of 137 published between 2005 and 2009, and none of 167 published between 1995 and 2004 (ref. 1). Accordingly, the use of OS as a primary end point decreased over time, from 49% in 1995–2004, to 36% in 2005–2009, and to 29% in 2010–2020. In another study, 67% of oncology drug approvals granted by the FDA between 2008 and 2012 were found to be based on a surrogate end point, 47% of which were based on PFS or DFS (31% overall)[4].

PFS is the most frequently used composite end point in oncology and is defined as the time from randomization to disease progression or death from any cause, whichever occurs first. Most commonly, progression of solid tumours is determined using the revised Response Evaluation Criteria In Solid Tumors (RECIST 1.1) guidelines[5], which define progressive disease as an increase of ≥20% in the sum of diameters of target lesions (maximum of five and two per organ, each >10 mm in longest diameter) and an absolute increase of ≥5 mm from their smallest sum diameters recorded on study (nadir value), the appearance of new lesions, or an unequivocal increase in non-target disease (that is, an increase in overall tumour burden, and not a single non-target lesion, sufficient to merit discontinuation of therapy). PFS is, therefore, a composite of at least five different events (Fig. 1): (1) death, (2) new lesions, (3) 20% growth of target lesions without any shrinkage, (4) shrinkage of target lesions followed by 20% growth, and (5) clear non-target disease growth, whichever happens first.

Other composite end points frequently used in oncology clinical trials include EFS, DFS and RFS. EFS is typically used in trials of neoadjuvant therapies and is defined as the time from randomization to disease recurrence and/or disease progression or death from any cause[6]. DFS is similarly defined as the time from treatment to disease recurrence or death from any cause, but is typically applied in the adjuvant setting after patients have received definitive, curative-intent local therapy (such as surgery and/or radiotherapy)[7]. Typically, no measurable disease (by conventional modalities) is present at the time of treatment initiation in trials using DFS, as opposed to neoadjuvant trials using EFS, in which the study treatment is initiated prior to definitive therapy. Similar to DFS, RFS is usually defined as the time from treatment to disease recurrence or death from any cause but, unlike most definitions of DFS, often excludes the development of second primary cancers[8].

Variability in component events of composite outcomes

The composite outcomes used in oncology are often vaguely defined, with definitions varying widely both between and within studies (Table 1). For many composite outcomes, no standardized guidelines exist and adherence to specific definitions is not mandated by regulatory agencies. Even when consensus definitions have been developed, they are not always uniformly adopted in clinical trial protocols[9]. Creating further confusion, composite end points that have different event compositions are often used interchangeably (for example, DFS and RFS or PFS and time to tumour progression (TTP), defined as the time from randomization to disease progression, excluding death).

Although the RECIST 1.1 guidelines have provided a consensus, unified definition of PFS that is applicable to most solid tumour types, they cannot be applied to all cancers. The definition of PFS is complicated by the fact that disease progression can be radiological, clinical and/or biochemical, depending on the study and setting. Consider the case of PFS in clinical trials involving patients with prostate cancer (Fig. 2). Historically, disease progression in this setting was defined biochemically based on an increasing serum level of prostate-specific antigen (PSA), which has been shown to be a poor surrogate for OS[10]. In 1999, the Prostate Cancer Clinical Trials Working Group (PCWG) introduced the first consensus definition of disease progression in patients with prostate cancer; these PCWG1 criteria include either an increase of ≥50% in serum PSA from the nadir level with an absolute increase of ≥5 ng/ml (or back to the baseline level, if lower), one or more new lesions on a bone scan or progression of nodal or parenchymal disease assessed radiographically or by physical examination[11]. In 2000, the original RECIST (1.0) guidelines were published[12] and subsequently became the gold standard for assessment of radiological progression in patients

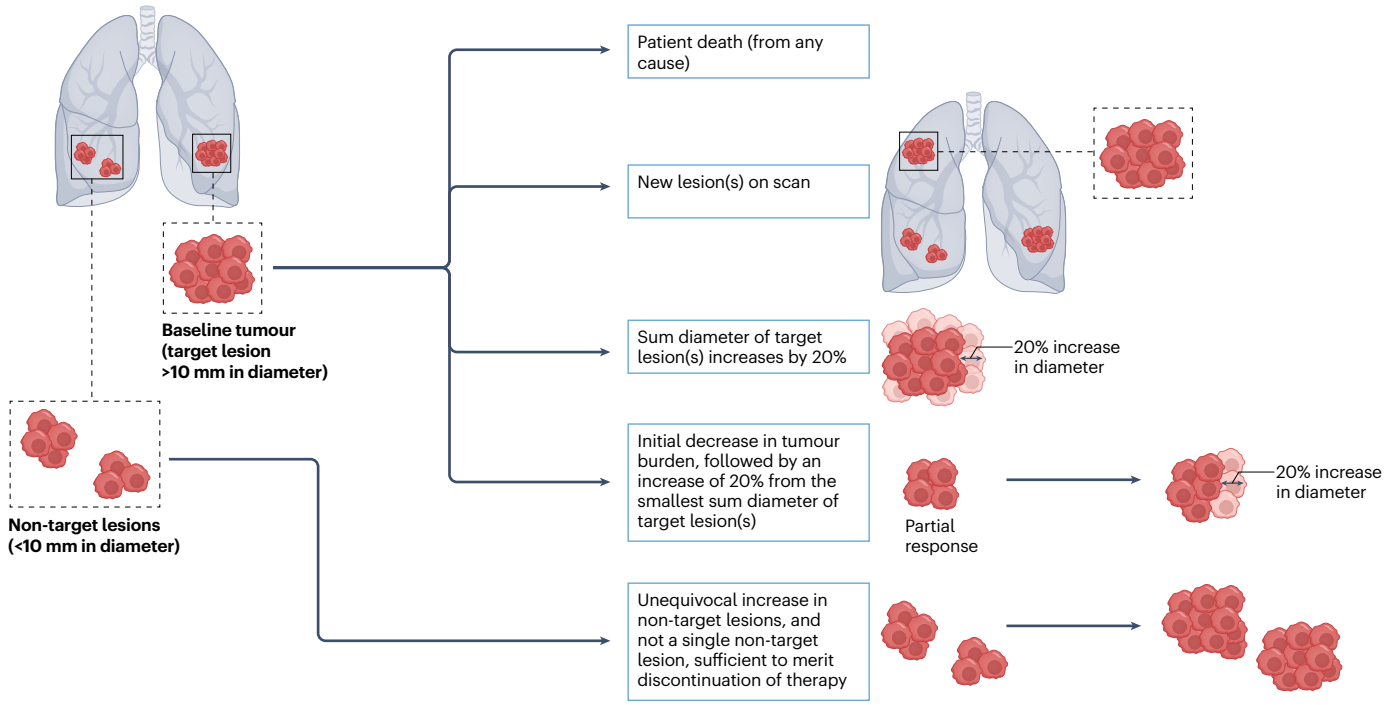

Fig. 1 | Components of PFS according to RECIST definitions for disease progression. The Response Evaluation Criteria In Solid Tumors version 1.1 (RECIST 1.1)[5] defines disease progression as either a 20% increase and 5 mm absolute increase in the smallest sum of the longest diameters of target lesions (maximum of five and two per organ, each >10 mm in longest diameter) recorded on study, the appearance of new lesions, or an unequivocal increase in non-target lesions. Therefore, progression-free survival (PFS) is a composite outcome comprising at least five different component events: (1) death, (2) new lesions, (3) 20% growth of target lesions without any shrinkage, (4) shrinkage of target lesions followed by 20% growth, and (5) a clear increase in the burden of non-target lesions, whichever happens first.

with solid tumours. However, RECIST 1.0 did not fully capture progression of prostate cancer, as these criteria excluded certain important aspects that are characteristic of this disease, such as PSA levels and bone lesions. In 2008, the PCWG proposed updated guidelines for response assessment of prostate cancer (PCWG2) that altered the thresholds for PSA progression to ≥25% and ≥2 ng/ml above the nadir and bone progression to two or more new lesions on bone scan, and also incorporated modified RECIST definitions of soft-tissue disease progression[13]. Whereas the PCWG1 criteria and RECIST focus on biochemical and/or radiological disease progression, the PCWG2 criteria also include symptomatic progression events (for example, worsening pain or HRQOL)[13]. Notably, however, no standardized definition of symptomatic progression was proposed. In 2016, the PCWG3 response assessment guidelines were published with further updated disease progression criteria, emphasizing the need to report various sites of progression independently and distinguish the appearance of new lesions from the growth of existing lesions[14]. Not all of these definitions have been fully accepted by clinical trialists; some studies in patients with prostate cancer use bespoke criteria or modified versions of the RECIST or PCWG2/3 criteria[15–18]. Other composite end points such as radiological PFS (rPFS) (radiographic progression or death) or PSA PFS (biochemical progression or death) are also widely used and are often defined using the relevant PCWG2 criteria for disease progression[19,20].

The heterogeneity of composite end point definitions makes interpreting trial results extremely challenging. Moreover, altering the component events included in a composite outcome can lead to drastically different conclusions. For example, one single-arm study of enzalutamide in patients with metastatic castration-resistant prostate cancer (mCRPC) demonstrated a median PSA PFS of 16 weeks but a median rPFS of 37.7 weeks — a more than twofold difference[21]. Other studies have produced similar results[22,23], suggesting that radiological and biochemical disease progression in patients with prostate cancer are distinct events with potentially different clinical and prognostic implications. Consequently, a composite outcome encompassing radiological, clinical and biochemical progression might be of very different value compared with a composite of radiological and clinical progression, or an end point capturing radiological or biochemical progression alone. Thus, comparing results from trials using different measures of PFS, which are likely to vary considerably in terms of surrogacy for OS, might lead to incorrect inferences about the relative clinical benefits of the treatments evaluated, which could potentially translate into suboptimal treatment decisions.

Indeed, a treatment might ultimately be deemed effective or not effective depending on whether clinical, biochemical and/or radiological definitions of progression are included as components of the primary end point. For example, the placebo-controlled phase III ACIS trial testing the addition of apalutamide to abiraterone and prednisone in patients with mCRPC demonstrated no statistically significant difference in time to PSA progression between the experimental and control arms (median 13.8 months versus 12.0 months; HR 0.87, 95% CI 0.74–1.02; $P = 0.076$) despite a significant improvement in the primary end point of rPFS (22.6 months versus 16.6 months; HR 0.69, 95%

Table 1 | Selected examples of composite end points and their definitions used in oncology clinical trials

End point	Trial	Cancer type	Outcomes included in composite
PFS	KEYNOTE-024 (ref. 99)	NSCLC	Disease progression per RECIST 1.1 (≥20% increase in smallest sum of longest diameters of target lesions and an absolute increase of ≥5 mm, one or more new lesions, or unequivocal increase in non-target lesions sufficient to warrant discontinuation of therapy)[5] by blinded independent central review, or death from any cause
PFS	CALGB 9182 (ref. 100)	Castration-resistant prostate cancer	Two or more new lesions on bone scan, increase of serum PSA >100% above pretreatment baseline level, worsening performance status ≥1, or death from any cause
rPFS	PREVAIL[101]	Metastatic prostate cancer	Disease progression per RECIST 1.1 for soft-tissue disease and PCWG2 for bone disease (two or more new lesions on bone scan)[13] by blinded independent central review, or death from any cause within 168 days of treatment discontinuation
PFS	SWOG S0777 (ref. 102)	Multiple myeloma	Increase of ≥25% from baseline in serum M component, urine M component, the difference between involved and uninvolved serum free light chain levels, or bone marrow plasma cell percentage; definitive development of or increasing size of bone lesions or soft-tissue plasmacytomas; development of hypercalcaemia solely due to myeloma; or death from any cause
DFS	KEYNOTE-091 (ref. 103)	NSCLC	Disease recurrence per RECIST 1.1 by investigator review, appearance of second primary malignancy, or death from any cause
DFS	PETACC-3 (ref. 104)	Colon cancer	Local, regional or distant relapse, second primary malignancy (colon or other), or death from any cause
iDFS	ExteNET[105]	HER2$^+$ breast cancer	Invasive ipsilateral tumour recurrence, invasive contralateral breast cancer, local or regional invasive recurrence, distant recurrence, or death from any cause
EFS	RATIFY[106]	Acute myeloid leukaemia	Disease relapse; failure to achieve complete remission, with complete remission defined as <5% blasts in bone marrow or extramedullary leukaemia, an absolute neutrophil count >1,000/μl, a platelet count >100,000/μl or an absence of blasts in peripheral blood; or death from any cause
EFS	CheckMate 816 (ref. 107)	NSCLC	Any progression of disease (per RECIST 1.1 by blinded independent central radiological review) precluding surgery, progression or recurrence of disease after surgery, progression of disease in the absence of surgery, or death from any cause
RFS	ZUMA-3 (ref. 108)	B cell-precursor acute lymphoblastic leukaemia	>5% blasts in bone marrow, circulating leukaemia present, CNS disease grade 2 or 3, progressive disease (defined as an increase of ≥50% from the nadir of the sum diameters of at least two lymph nodes or one if only a single node is involved, an increase of ≥50% in the longest diameter of any single previously identified node >1 cm in its short axis, or an increase of ≥50% in the size of splenic, hepatic or any other non-nodal lesion), or death from any cause
RFS	KEYNOTE-716 (ref. 109)	Melanoma	Disease recurrence as assessed by investigator (via physical examination with biopsy confirmation, or radiographically), or death from any cause

CNS, central nervous system; DFS, disease-free survival; EFS, event-free survival; iDFS, invasive disease-free survival; NSCLC, non-small-cell lung cancer; PCWG2, Prostate Cancer Working Group 2; PFS, progression-free survival; PSA, prostate-specific antigen; RECIST 1.1, Response Evaluation Criteria In Solid Tumors version 1.1; RFS, relapse-free or recurrence-free survival; rPFS, radiological progression-free survival.

CI 0.58–0.83; $P < 0.0001$)[24]. Furthermore, a post hoc analysis involving patients with mCRPC receiving abiraterone and prednisone in the phase III COU-AA-302 trial demonstrated that the inclusion of unequivocal clinical progression events (defined as treatment discontinuation owing to a worsening of clinical status) in addition to radiographic progression and death as a component of PFS lowered estimates of treatment benefit compared with rPFS alone (median clinicoradiographic PFS of 13.3 months versus median rPFS of 16.5 months)[25]. The choice of composite events to include in the primary end point of a trial is, therefore, crucial and can influence conclusions about the therapeutic efficacy and clinical benefit, ultimately affecting treatment decisions.

Heterogeneity of PFS definitions can also complicate integration of results from different studies in meta-analyses, especially if treatment effects vary by component event. Some researchers have suggested that meta-analyses of trials using composite end points should be performed by individual components rather than by composite outcome[26,27]. Nevertheless, trials with varying definitions of PFS are often combined in meta-analyses. For example, in a meta-analysis of RCTs evaluating docetaxel-based chemotherapy with or without anti-angiogenic agents for the treatment of mCRPC, the five RCTs included in the PFS analysis used different definitions of progression[28]. One trial defined PFS using the PCWG1 criteria modified to require two or more new bone lesions for bone scan progression[29]. A second trial defined disease progression based on soft-tissue disease according to RECIST, new lesions on bone scans, or increasing pain and/or analgesia, but excluded PSA progression alone[30]. Another trial defined progression as four or more new bone lesions, increased pain, new or progressive soft-tissue disease according to modified RECIST, or skeletal-related events such as pathological fracture or vertebral compression[18]. Yet another study did not use PFS, but rather TTP, and did not specify criteria for disease progression[31]. In the final study, PFS was based on radiographic or clinical disease progression but no further detail was provided[32].

Variation in radiographic response criteria

Deviations from the RECIST definitions are also frequent in PFS end points of studies focused on other tumour types beyond prostate cancer. For example, both the Response Evaluation Criteria In Lymphoma (RECIL)[33] and the Lugano criteria[34] are commonly used for response assessment in patients with lymphoma, and these two systems have shown discordance for the classification of progressive disease — in one

study, 85.7% of patients with progressive metabolic disease according to the Lugano criteria had a partial or minimal response by RECIL[35]. The Lugano criteria rely primarily on [^{18}F]fluorodeoxyglucose (FDG) uptake on PET–CT, assessed using Deauville scores, for the evaluation of up to six target lesions, although bidimensional perpendicular diameter measurements on CT can be used for non-FDG-avid tumours[36]. The RECIL definitions of response also incorporate Deauville scores but rely on unidimensional measurements of the diameters of up to three lesions, and CT measurements alone are used to define progressive disease[33]. Other radiographic response evaluation systems used for PFS assessment in patients with particular tumour types include the Response Assessment in Neuro-Oncology (RANO) criteria for central nervous system (CNS) tumours[37], the Choi criteria for gastrointestinal stromal tumours[38] and the modified RECIST (mRECIST) for hepatocellular carcinoma[39]. The RANO group have also proposed response criteria specifically for brain metastases (RANO-BM)[40], as opposed to primary CNS tumours, adding further complexity to PFS assessments when separate systems are used for CNS and non-CNS disease. Moreover, multiple radiographic response guidelines, including irRC[41], irRECIST[42] and iRECIST[43], have been developed with the specific aim of better accounting for the atypical patterns of tumour response that can occur with immunotherapies, such as pseudoprogression (an initial increase in tumour size or the appearance of new lesions, owing to increased immune cell infiltration into tumours, followed by a decrease in tumour burden). The most widely adopted immunotherapy response criteria, iRECIST, are similar to RECIST 1.1 but require confirmation of progressive disease with follow-up imaging; the levels of discordance between iRECIST and RECIST 1.1 vary depending on the study[44,45].

Even within RECIST, the choice of imaging tool — CT or MRI — can complicate PFS measurements; MRI, the less preferred choice, has higher sensitivity for certain lesions but is also prone to artefacts[46]. Researchers should be aware of the variation between and within composite end point definitions given that the criteria used might affect the correlation of PFS as a surrogate for OS, raising questions about the true benefit of treatment. In addition, variation in definitions can limit cross-trial comparisons even when other factors, including inclusion and exclusion criteria, are well matched.

Heterogeneity in EFS and DFS end points

Other composite end points such as EFS and DFS are even more heterogeneously defined than PFS owing to lack of consensus definitions. In studies focused on leukaemias, for example, EFS has been defined in at least 12 different ways in RCTs[47]. The component events included in these definitions vary widely, ranging from only disease recurrence and death[48] to induction failure, disease relapse, secondary malignancy and death[49]. Some trials also consider discontinuation of therapy owing to toxicity as an EFS event[50]. Studies also vary in whether loss of haematological or cytogenetic response is included

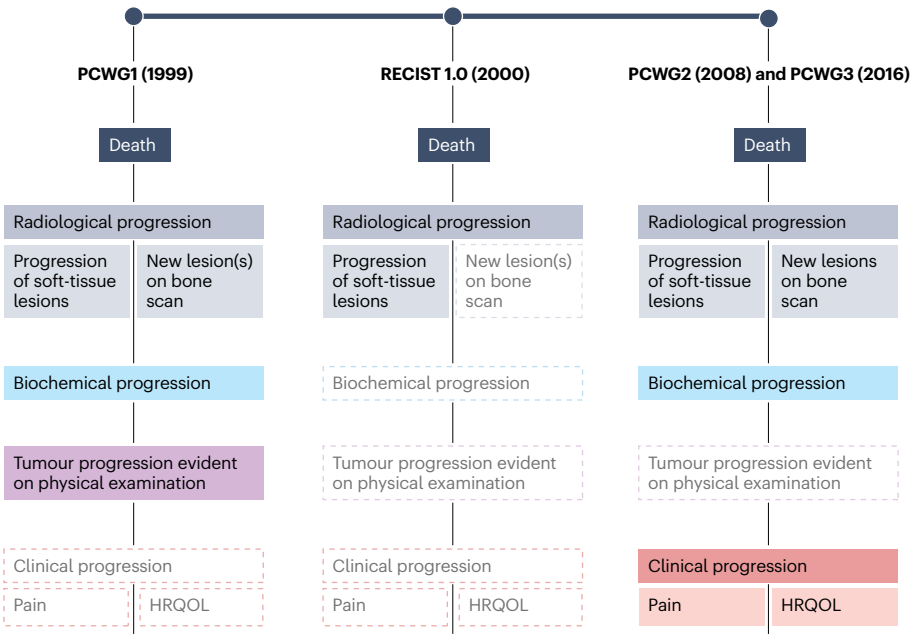

Fig. 2 | Definitions of PFS in patients with prostate cancer according to standard response criteria. The composite end point of progression-free survival (PFS) consists of death from any cause as well as cancer progression, which in the setting of prostate cancer has been defined variably according to heterogeneous response criteria. The PCWG1 consensus response criteria proposed by the Prostate Cancer Working Group (PCWG) in 1999 defined progression of prostate cancer either biochemically as an increase of ≥50% in serum prostate-specific antigen (PSA) from the nadir level with an absolute increase of ≥5 ng/ml (or back to the baseline level, if lower), radiologically as one or more new lesions on a bone scan or progression of soft-tissue disease on imaging, with evidence from physical examination also applicable for the latter[11]. In 2000, the original Response Evaluation Criteria In Solid Tumors (RECIST 1.0) were published and defined disease progression based only on radiological assessment of soft-tissue lesions as an increase of 20% in the smallest sum of longest diameters of target lesions, the appearance of new lesions, or an unequivocal increase in non-target lesions[12]. Subsequently, the PCWG2 (ref. 13) and PCWG3 (ref. 14) were proposed to incorporate the RECIST definitions for radiological progression of soft-tissue lesions as well as to revise the criteria for progression based on bone scans to require two or more new lesions, and also included new clinical criteria for capturing symptomatic progression based, for example, on worsening pain or health-related quality of life (HRQOL). In addition, the definition of biochemical progression was altered to an increase in serum PSA of ≥25% and ≥2 ng/ml above the nadir.

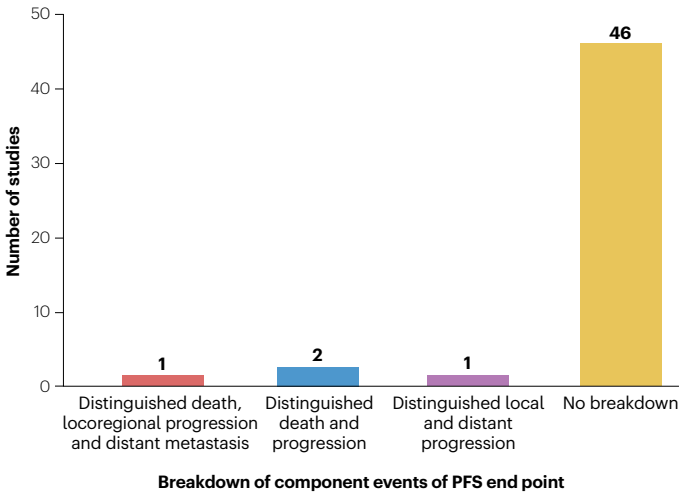

Fig. 3 | Evaluation of the reporting of PFS component events in oncology studies. We reviewed a sample of 50 studies using progression-free survival (PFS) as an outcome, encompassing the ten most recently published studies with results published in each of five top oncology journals publishing clinical research (according to Google Scholar metrics[61]) (Supplementary Table 1). We found that only three studies (6%) reported the number of deaths and progression events separately, and only one of these studies distinguished locoregional progression from distant metastasis. In addition, one study distinguished local and distant progression but did not report the number of deaths counted as PFS events.

in EFS definitions[51,52]. In studies that use failure to achieve a complete response (CR) as a measure of induction failure in the EFS analysis, the timing of CR assessment is also important. Yin et al.[53] evaluated EFS as an end point using data from five Cancer and Leukaemia Group B (CALGB) trials involving patients with previously untreated acute myeloid leukaemia and three different definitions of induction failure: failure to achieve a CR by 60 days after patient randomization or registration, by the end of all courses of induction therapy, and by the end of all protocol-defined treatment. They found that median EFS estimates for individual trials varied by 14% to 115% depending on the definition of induction failure applied to the data[53]. Therefore, evaluating therapeutic efficacy based on single-arm trials using EFS results alone might result in incorrect conclusions.

The inclusion of second primary cancers in DFS definitions has been shown to affect study conclusions. In the phase III PETACC-3 trial involving patients with colon cancer, the primary DFS end point was not met at the 3-year follow-up assessment when second primary cancer was included as a component event (as per the definition in the protocol); however, a statistically significant improvement in RFS, which excluded second non-colon primary cancers, was observed in the experimental arm[54,55]. Tolaney et al.[56] performed a simulation modelling study to examine the effect of including second non-breast primary cancers in invasive DFS (iDFS) definitions in trials of adjuvant therapy for breast cancer and found that inclusion of such events can result in false conclusions of non-inferiority, even if the occurrence rates are low and similar between standard-of-care and the experimental arms.

To address the variation in DFS definitions across breast cancer clinical trials, the STEEP criteria provided a formal definition of iDFS in 2007 (ref. 57). However, an analysis of phase III RCTs included in the aforementioned study by Tolaney et al.[56] revealed that only three (27%) of 11 studies complied with the STEEP criteria; three trials excluded second non-breast primary cancers from iDFS definitions, and three others used appropriate criteria but erroneously referred to iDFS as DFS. Confusion between similarly defined composites is not uncommon. For example, another study of trials involving patients with breast cancer revealed that definitions of TTP often include death (13 of 16 trials with TTP as a primary end point, excluding five in which no definitions were reported), making PFS the most appropriate term[58].

Effects on surrogacy analyses

The inconsistent definitions of composite outcomes complicate surrogacy analyses, which are a type of meta-analysis used to evaluate the correlation between a surrogate end point and OS across multiple RCTs. Such trial-level analyses aim to determine whether treatments that improve a surrogate end point also improve a clinically meaningful end point[59]. Considered the gold standard for determining the validity of a surrogate, surrogacy analyses should be robust, as their conclusions can influence future drug approval decisions. Ajani et al.[60] conducted a surrogacy analysis of DFS in trials involving patients with resectable oesophageal or gastro-oesophageal junction cancer. The authors initially defined DFS as time from randomization to either disease recurrence or death, but found that few RCTs used this definition. As a result, the surrogacy analysis was broadened to include alternative definitions of DFS and PFS that varied considerably between trials — for example, the starting time for DFS ranged from the day of surgery to 1 week or up to 6 months after surgery[60]. On the basis of their analysis of 26 trials, the authors concluded that DFS — defined as time from tumour resection to disease recurrence or death — is a valid and useful surrogate for OS in the neoadjuvant, perioperative or adjuvant settings; however the degree to which the heterogeneity in DFS definitions as a possible source of bias affected this conclusion is unknown.

Lack of transparent reporting of composite events

In cardiology trials reporting MACE, common practice is to provide the frequency of the individual component events, such as stroke, myocardial infarction, hospitalization and death, alongside the composite outcome. This standard is not, however, the case for PFS and other composite end points of clinical trials in oncology. We analysed 50 studies using PFS with results published in five top oncology journals, according to Google Scholar metrics[61] (Fig. 3; Supplementary Table 1). For each of the journals, the ten most recently published research studies using PFS as an outcome measure, excluding meta-analyses, were selected. We found that only three studies (6%) reported the number of deaths versus disease progression events; only one such study distinguished locoregional progression from distant metastasis (Fig. 3). Additionally, one study distinguished local and distant progression but did not provide the number of deaths prior to disease progression.

The reasons behind the difference in composite reporting standards between cardiology and oncology are unclear. One possibility is that composites such as PFS and DFS are efficacy outcomes. MACE, however, originated as a safety outcome and was first used in studies focused on complications of percutaneous coronary interventions[62]. Regulatory agency standards for reporting safety outcomes are high, and thus detailed documentation of adverse events in clinical trials is required. By the time MACE became widely used as an efficacy end point, providing numbers of each event might have been convention. Another possibility underlying the difference in reporting standards is that PFS is considered to be a collection of similar events, whereas MACE is considered to be a collection of dissimilar events (for example,

stroke versus myocardial infarction). Yet, primary tumour growth and distant metastasis are markedly different, particularly in terms of their clinical implications. Finally, historical accident — that is, PFS originating as an extension of TTP — might explain the different reporting standards. Ultimately, all these explanations are speculatory, but notably none justifies incomplete reporting. We contend that transparent reporting of component events must become the norm for oncology trials using composite end points, particularly if the composite is a primary end point or is used for regulatory purposes, but also when presented as a secondary end point.

Composite events vary in clinical significance and frequency

Even focusing on a single composite end point such as PFS, the component outcomes of primary tumour and/or target lesion growth, the appearance of new lesions or metastases, and death are not equally clinically significant. Metastasis rather than primary tumour growth is the predominant determinant of a poor prognosis, with 90% of cancer-related deaths occurring owing to metastatic disease[63]. Indeed, the use of other composite end points such as metastasis-free survival[64], locoregional relapse-free survival[65] and distant metastasis-free survival[66] suggest that the extent of metastatic spread is important. Moreover, the prognosis of patients with distant metastases is often poorer than of those with locoregional lymph node metastases[67], and might vary further depending on the organ involved[68].

Multiple studies have shown that the type of progressive disease according to RECIST criteria affects prognosis. Twelves et al.[69] and Mori et al.[70] found that patients with metastatic breast cancer who had disease progression attributed to an increase in the size of pre-existing lesions had better survival than those who had progression owing to the appearance of new lesions. Similar findings have been reported in an analysis of patients with metastatic colon cancer who had a best response of progressive disease (according to RECIST 1.0) in the phase III Nordic VI trial[71]. Without the development of a new lesion or progression of a non-target lesion, an increase in target lesion diameter of ≥20% was not associated with a statistically significant OS detriment compared with an increase of <10%. In addition, an analysis of stratified PFS data from the phase III RECORD-1 trial in patients with metastatic renal cell carcinoma revealed that growth of a non-target lesion and/or the appearance of a new lesion at the first assessment after the baseline assessment (performed after 2–14 weeks) was predictive of OS (univariate $P < 0.001$), whereas a change in the sum of tumour diameters of −30% to +10% was not[72].

Distinguishing between various progression events might be crucial for the proper interpretation of PFS results: if an improvement in PFS is mostly attributable to a reduced incidence of less clinically significant events, the true clinical benefit to the patient will remain unclear. In situations of discordancy between PFS and OS — that is, PFS improvements that do not translate into OS benefits — understanding the breakdown of PFS events and comparisons across arms might be key to making sense of the results.

Indeed, the different clinical significance of distinct component events might explain why composite outcomes are often poor surrogates for OS. An empirical analysis of 78 surrogate outcome validation studies in oncology revealed that only 12% of surrogate indication pairs had a strong correlation with OS, and 38% had a weak correlation[73]. For example, in patients with hormone receptor-positive breast cancer, extended adjuvant endocrine therapy has not consistently demonstrated OS improvements despite DFS benefits[74]. Within DFS, some events such as locoregional recurrences are more likely to be amenable to surgical resection (and possible cure) and thus less likely to affect OS than distant recurrences. An analysis of six trials of adjuvant endocrine therapy for breast cancer including 23,371 patients revealed that the proportion of contralateral breast cancer-related and non-breast cancer-related deaths increased with follow-up duration, while the proportion of distant and locoregional recurrence decreased[75]. The authors of this study concluded that improvements in DFS with adjuvant therapy might not translate into an OS benefit, as adjuvant therapy might predominantly affect the frequency of more curable events. Another analysis of 84 trials of adjuvant therapy for breast cancer showed that the proportion of randomized patients who have distant recurrence or other breast cancer-related events has declined over the past two decades[76]. Importantly, almost one third of 165 trials initially identified by the authors were excluded from their analysis owing to lack of information on the breakdown of DFS events[76]. In these trials, determining whether the adjuvant therapy tested would result in any clinically meaningful benefit to the patient is essentially impossible based on DFS results alone.

The results of multiple surveys indicate that the majority of patients are not willing to receive therapy that delays radiological disease progression but with additional toxicities and no concomitant benefits in OS or HRQOL[77,78]. Notably, this consideration is likely to be even more pertinent when the frequency of clinically significant composite events is substantially lower than that of less significant events, or if imbalances in these ratios exist between study arms, given that any improvement in a composite outcome is less likely to translate into a clinically meaningful OS or HRQOL benefit in these scenarios[3].

Component events can be differentially affected by therapy

Solely reporting a composite outcome might also lead to assumptions that the treatment effects apply to each of the component events, which might not necessarily be true. At times, study authors might even encourage such misinterpretation. For example, a systematic review revealed that in the majority (69%) of publications reporting RCTs with a statistically significant improvement in a composite end point, the abstracts falsely implied that the treatment effect applied to the most important component, as determined by two independent, blinded observers[79].

To avoid this problem, the EMA guidelines for evaluating anticancer therapies in patients recommend reporting "separate analyses for individual types of events using descriptive summary tables and, where appropriate, competing-risks approaches to explore treatment effect on the various types of events"[80]. In practice, however, many studies are not adequately powered to detect significant differences in the individual components of a composite outcome[81].

Thus, the effects of treatment on a composite end point cannot be used to infer effects on the individual component events. A therapy that reduces the size of pre-existing tumours, for example, might not actually decrease the patients' probability of developing new lesions. The relationships between tumour growth, metastasis and mortality are highly complex and not fully elucidated. Increasing evidence indicates that disease progression is a less intuitive process than classic models have implied. Whereas distant metastasis was traditionally believed to result from a linear process of cancer evolution, emerging data suggest that metastatic disease can develop in parallel to the primary tumour in some cancers. For example, a study in patients with invasive breast cancer revealed a non-linear correlation between

tumour size and lymph node metastasis[82]. For very small tumours (<10 mm in diameter) or for very large tumours (>60–90 mm), rates of lymph node metastasis and breast cancer mortality remained relatively constant as tumour size increased. For instance, the rate of lymph node involvement was approximately equal for 70-mm tumours and 150-mm tumours. According to this parallel model of cancer evolution, the primary tumour might not be the source of all metastases, and thus therapy directed towards the primary tumour might not affect metastatic disease owing to the distinct biology of lesions arising through parallel evolution[83]. Potentially heterogeneous treatment effects relating to these complexities of cancer progression might be misunderstood if complete data on the different component events of a composite outcome are not provided.

The heterogeneity of treatment effects is affected by the underlying biology of the tumours and mechanism of action of the therapy. In patients with prostate cancer, serum PSA levels are prone to fluctuation, which might explain why changes in PSA levels are less indicative of treatment efficacy than objective radiological responses[19]. Moreover, disease progression on imaging despite stable or declining PSA levels has been frequently documented[84], suggesting that radiographic and biochemical progression reflect distinct biological processes. Furthermore, several clinical trials of immune-checkpoint inhibitors and certain other immunotherapies have demonstrated OS benefits without improvements in PFS[85]. Part of this disconnect might reflect a potential ability of these agents to affect the more OS-limiting components of PFS (for example, the development of new lesions to a greater extent than the growth of existing lesions) owing to their complex mechanisms of action that are not limited to direct cytotoxicity in measurable tumour masses[85]. Interpreting treatment effects becomes more complicated when symptomatic progression is included in PFS definitions, considering that alterations in patient-reported outcomes such as pain scores might be more reflective of hospital-based care than the direct effects of an experimental therapy. Ideally, each component of a composite outcome should be representative of the same causal biological pathway. The more inconsistent the treatment effects between distinctly different clinical events, the more challenging it is to accurately interpret a PFS result.

A key assumption underlying the use of a composite end point is that the components will be altered by treatment in a similar way. Occasionally, however, treatment effects on variables such as primary tumour growth, metastasis and mortality can paradoxically occur in opposite directions. For example, a highly toxic drug might decrease tumour spread but increase mortality. This situation occurred in the BELLINI trial involving patients with relapsed and/or refractory multiple myeloma, which demonstrated improved PFS but lower OS in the experimental arm owing to a higher rate of infections related to therapy[86]. In addition, some preclinical data suggest that shrinkage of a primary tumour using chemotherapy can accelerate distant spread in some contexts by selecting for stem-like cells that are more likely to initiate metastasis[87]. The directionality assumption is less likely to hold true when a greater number of events are included in the composite end point, as is the case for certain definitions of PFS, EFS and DFS. Allogeneic haematopoietic stem cell transplantation (alloHSCT) is a common situation in which opposite treatment effects can occur. Trials of treatments involving alloHSCT often use a composite end point comprising death, cancer relapse and graft versus host disease (GVHD), referred to as GVHD-free RFS (GRFS). In this context, a therapy that reduces the incidence of GVHD might increase the rate of cancer relapse (or, vice versa, a treatment that reduces relapse rates might result in more GVHD); thus, the directionality assumption is often violated and interpretation of the clinical significance of the GRFS result is difficult[88,89]. In such cases, GVHD and relapse rates must be analysed individually to accurately gauge the risk to benefit ratio of treatment.

Routinely reporting the frequencies of the different events included in a composite outcome can provide assurance that treatment effects on individual components are in the same direction; any 'qualitative heterogeneity' (that is, difference in directionality) would render use of the composite outcome invalid[90].

Varying levels of bias among composite events

The components of a composite end point have varying levels of uncertainty. Among PFS and DFS events, time to tumour growth or metastasis is subject to measurement bias given that the results are dependent on the assessment schedule, skill of the radiologist, choice of imaging modality and whether the investigator is blinded to treatment allocation. Whereas time to death is usually known precisely, time to progression is intermittently assessed and assigned only after progressive disease is detected, typically resulting in overestimation of PFS[91]. The amount of bias in a PFS measurement increases with the ratio of progression to death events[91]. Providing a full breakdown of events can therefore provide information on the likely extent of bias in a composite outcome.

The ratio of events varies by tumour and treatment type

Knowledge of the precise mixture of composite events is necessary to compare study populations, considering that ratios of events probably vary across different cancer and treatment types, although the available data are limited owing to lack of uniform reporting. For example, few DFS events occurring in patients with localized hormone receptor-positive breast cancer would be expected to be deaths, whereas this might not be the case for PFS events among patients with advanced-stage pancreatic cancer.

Even within a single cancer type and treatment setting, the relative effects of a therapy on different sites of disease can vary. For example, extended follow-up data from the ADAURA trial of adjuvant osimertinib in patients with resectable stage IB–IIIA *EGFR*-mutant NSCLC demonstrate that this agent clearly reduced the risk of DFS events (4-year DFS of 73% versus 38% with placebo); however, the magnitude of the reduction varied by site of disease recurrence[92]. The lymph node recurrence rate decreased from 17% with placebo to 5% with osimertinib, a reduction of >70%, whereas the rate of CNS recurrence decreased from 11% to 6.4%, a 42% reduction. Whether these differences occurred owing to chance variation or reflect a unmet need for agents with even greater CNS penetrance than osimertinib remains unknown; yet, such considerations can only be explored with transparent data reporting, as in this example.

As discussed, systemic therapies might have different effects on the various components of composite end points. Some treatments might primarily reduce target lesion tumour growth but not affect the rate of metastasis, whereas others might variably reduce the growth and/or incidence of metastatic lesions in different organs, and some might affect all components proportionately. Ideally, therefore, relative consistency in the various components of a composite outcome should be ensured before results from multiple studies are combined in meta-analyses.

Advantages of composite end points

The primary reason for using composite end points is to increase the statistical power of a study. Owing to the higher number of events,

the use of composite outcomes can reduce the sample size, costs and follow-up durations of clinical trials, which can expedite drug approvals. In a retrospective study encompassing 107 oncology drugs with 188 approved indications, the use of PFS as a surrogate end point in registration trials reduced the mean time needed to obtain the data supporting approval by 11 months compared with the use of OS as the primary end point; the authors estimated that this reduction would expedite the typical overall clinical development time of a drug by approximately 12%[93]. In addition, some events, such as second malignancies or certain adverse events, are too rare for individual analyses to be feasibly conducted.

Notably, however, the effects of composite end points on statistical efficiency have a number of caveats. For example, statistical power can be decreased if treatment effects on different components of the composite outcome occur in opposite directions. Moreover, even if the effects are in the same direction, the sample size requirement can still be considerable if a high degree of overlap exists between the individual components of a binary composite end point and the relative treatment effects[94]. In addition, concerns have been raised that the use of composite outcomes might cause trialists to limit sample sizes (even when larger sample sizes are possible), given that decisions on sample size are typically based on rates of the composite outcome in similar trials rather than the frequencies of individual component events[95]. This approach can lower the internal and external validity of a study[96] and reduce the ability to perform component-level analysis.

Other advantages of composite end points include the assessment of treatment effects on outcomes beyond OS, which can be particularly useful when multiple aspects of a malignancy have similar clinical significance. Composite outcomes can also capture the net effect of treatment as long as the different component events correspond to the same response pathway. Furthermore, composite time-to-event end points often avoid the problem of competing risks that can complicate survival analyses. For example, in older patients and/or those with multiple morbidities or indolent cancers, death due to a non-cancer-related cause might precede death due to disease progression; therefore, an outcome focused on disease progression alone might wrongly suggest clinical benefit from a treatment that raises the risk of non-cancer-related death.

Conclusions

The use of composite outcomes as primary end points in oncology trials has steadily gained popularity over the past three decades. Although these end points have made RCTs more efficient and have modestly accelerated drug approvals, they are prone to bias and misinterpretation. Therapeutic effects on composite outcomes alone might not always faithfully guide clinical decision-making.

Composite end points such as PFS, DFS, EFS and RFS are often inconsistently or imprecisely defined; the particular definition used can greatly affect study conclusions and, in some cases, determine whether or not treatment effects are statistically significant. This inconsistency complicates the comparison of results across multiple trials, and heterogeneously defined composite outcomes are often analysed together in meta-analyses. Several groups have made efforts to address these issues. For example, the DATECAN initiative seeks to define time-to-event end points for oncology RCTs and has already provided international consensus recommendations for such end points in trials focused on pancreatic cancer[97], sarcomas and gastrointestinal stromal tumour[65], and renal cell cancer[98]. However, even when standardized definitions do exist (for example, the STEEP criteria for iDFS in the setting of breast cancer), they are not always adopted by clinical trialists. Hence, regulatory agencies must increase efforts to mandate adherence to standard criteria in RCTs.

For a composite end point to be valid, the component events should generally be of similar clinical significance, occur at comparable frequencies and be similarly affected by treatment. Transparent reporting of component events in oncology trials is not the norm but is essential to ensure that these criteria are met and that use of the composite end point is appropriate. Without a complete breakdown of component events, it is impossible to know whether an improvement in a composite outcome is driven by the most clinically significant events, thus leaving the true therapeutic benefit open to question. Including a full breakdown of a composite outcome also helps to avoid misinterpretation of trial results and provides a more nuanced understanding of treatment effects, which can vary greatly between component events. Furthermore, clear reporting of individual events is necessary to appropriately combine results from multiple clinical trials and compare patient populations.

Future work can explore three key avenues. First, in individual RCTs, reporting a full breakdown of composite outcomes might reveal imbalances in effects on different component events and thereby provide a clearer understanding of the effectiveness of treatments tested. Second, within a given cancer type, a portfolio of RCTs could be examined and individual outliers (that is, those with different ratios of component events) should be scrutinized. Third, the precise balance of component events can be analysed across cancer types and treatment settings to elucidate differences in the severity of illness and specific challenges faced by patients. Ultimately, such transparency and knowledge could facilitate treatment decisions by clarifying the benefits of a given therapy and might offer additional, alternative benefits not yet realized.

Published online: 12 October 2023

References

1. Del Paggio, J. C. et al. Evolution of the randomized clinical trial in the era of precision oncology. *JAMA Oncol.* **7**, 728–734 (2021).
2. Le-Rademacher, J. & Wang, X. Time-to-event data: an overview and analysis considerations. *J. Thorac. Oncol.* **16**, 1067–1074 (2021).
3. McCoy, C. E. Understanding the use of composite endpoints in clinical trials. *West J. Emerg. Med.* **19**, 631–634 (2018).
4. Kim, C. & Prasad, V. Cancer drugs approved on the basis of a surrogate end point and subsequent overall survival: an analysis of 5 years of US Food and Drug Administration approvals. *JAMA Intern. Med.* **175**, 1992–1994 (2015).
5. Eisenhauer, E. A. et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur. J. Cancer* **45**, 228–247 (2009).
6. Delgado, A. & Guddati, A. K. Clinical endpoints in oncology–a primer. *Am. J. Cancer Res.* **11**, 1121–1131 (2021).
7. Robinson, A. G., Booth, C. M. & Eisenhauer, E. A. Disease-free survival as an end-point in the treatment of solid tumours – perspectives from clinical trials and clinical practice. *Eur. J. Cancer* **50**, 2298–2302 (2014).
8. Cohen, R. et al. Guidelines for time-to-event end-point definitions in adjuvant randomised trials for patients with localised colon cancer: results of the DATECAN initiative. *Eur. J. Cancer* **130**, 63–71 (2020).
9. Gourgou-Bourgade, S. et al. Guidelines for time-to-event end point definitions in breast cancer trials: results of the DATECAN initiative (Definition for the Assessment of Time-to-event Endpoints in CANcer trials). *Ann. Oncol.* **26**, 873–879 (2015).
10. Fleming, M. T., Morris, M. J., Heller, G. & Scher, H. I. Post-therapy changes in PSA as an outcome measure in prostate cancer clinical trials. *Nat. Rev. Clin. Oncol.* **3**, 658–667 (2006).
11. Bubley, G. et al. Eligibility and response guidelines for phase II clinical trials in androgen-independent prostate cancer: recommendations from the Prostate-Specific Antigen Working Group. *J. Clin. Oncol.* **17**, 3461–3467 (1999).
12. Therasse, P. et al. New guidelines to evaluate the response to treatment in solid tumors. *J. Natl Cancer Inst.* **92**, 205–216 (2000).
13. Scher, H. I. et al. Design and end points of clinical trials for patients with progressive prostate cancer and castrate levels of testosterone: recommendations of the Prostate Cancer Clinical Trials Working Group. *J. Clin. Oncol.* **26**, 1148–1159 (2008).

14. Scher, H. I. et al. Trial design and objectives for castration-resistant prostate cancer: updated recommendations from the Prostate Cancer Clinical Trials Working Group 3. *J. Clin. Oncol.* **34**, 1402–1418 (2016).
15. Ryan, C. J. et al. Abiraterone in metastatic prostate cancer without previous chemotherapy. *N. Engl. J. Med.* **368**, 138–148 (2013).
16. Franzese, C. et al. The efficacy of stereotactic body radiation therapy and the impact of systemic treatments in oligometastatic patients from prostate cancer. *Cancer Med.* **7**, 4379–4386 (2018).
17. Gafita, A. et al. Early prostate-specific antigen changes and clinical outcome after ^{177}Lu-PSMA radionuclide treatment in patients with metastatic castration-resistant prostate cancer. *J. Nucl. Med.* **61**, 1476–1483 (2020).
18. Fizazi, K. et al. Phase III, randomized, placebo-controlled study of docetaxel in combination with zibotentan in patients with metastatic castration-resistant prostate cancer. *J. clin. oncol.* **31**, 1740–1747 (2013).
19. Woo, S. et al. Correlation between imaging-based intermediate endpoints and overall survival in men with metastatic castration-resistant prostate cancer: analysis of 28 randomized trials using the Prostate Cancer Clinical Trials Working Group (PCWG2) criteria in 16,511 patients. *Clin. Genitourin. Cancer* **20**, 69–79 (2022).
20. Shore, N. D. et al. Prostate-specific antigen (PSA) progression-free survival (PFS): a comparison of degarelix versus leuprolide in patients with prostate cancer [abstract]. *J. Clin. Oncol.* **29** (Suppl. 7), 12 (2011).
21. Kato, H. et al. Consequences of an early PSA response to enzalutamide treatment for Japanese patients with metastatic castration-resistant prostate cancer. *Anticancer Res.* **36**, 6141–6150 (2016).
22. Zhao, J. et al. AKR1C3 expression in primary lesion rebiopsy at the time of metastatic castration-resistant prostate cancer is strongly associated with poor efficacy of abiraterone as a first-line therapy. *Prostate* **79**, 1553–1562 (2019).
23. Kim, W. et al. Sequential use of the androgen synthesis inhibitors ketoconazole and abiraterone acetate in castration-resistant prostate cancer and the predictive value of circulating androgens. *Clin. Cancer Res.* **20**, 6269–6276 (2014).
24. Saad, F. et al. Apalutamide plus abiraterone acetate and prednisone versus placebo plus abiraterone and prednisone in metastatic, castration-resistant prostate cancer (ACIS): a randomised, placebo-controlled, double-blind, multinational, phase 3 study. *Lancet Oncol.* **22**, 1541–1559 (2021).
25. Rao, A. et al. Impact of clinical versus radiographic progression on clinical outcomes in metastatic castration-resistant prostate cancer. *ESMO Open* **5**, e000943 (2020).
26. Choi, S. W. & Cheung, C. W. The case of the misleading composite – one outcome is better than two. *Anaesthesia* **71**, 1101–1103 (2016).
27. Freemantle, N., Calvert, M., Wood, J., Eastaugh, J. & Griffin, C. Composite outcomes in randomized trials: greater precision but with greater uncertainty? *JAMA* **289**, 2554–2559 (2003).
28. Lei, N. et al. Docetaxel-based therapy with and without antiangiogenic agents as first-line chemotherapy for castration-resistant prostate cancer: a meta-analysis of nine randomized controlled trials. *Mol. Clin. Oncol.* **2**, 1182–1188 (2014).
29. Kelly, W. K. et al. Randomized, double-blind, placebo-controlled phase III trial comparing docetaxel and prednisone with or without bevacizumab in men with metastatic castration-resistant prostate cancer: CALGB 90401. *J. Clin. Oncol.* **30**, 1534–1540 (2012).
30. Quinn, D. I. et al. Docetaxel and atrasentan versus docetaxel and placebo for men with advanced castration-resistant prostate cancer (SWOG S0421): a randomised phase 3 trial. *Lancet Oncol.* **14**, 893–900 (2013).
31. Pili, R. et al. Phase II study on the addition of ASA404 (vadimezan; 5,6-dimethylxanthenone-4-acetic acid) to docetaxel in CRMPC. *Clin. Cancer Res.* **16**, 2906–2914 (2010).
32. Heidenreich, A. et al. A randomized, double-blind, multicenter, phase 2 study of a human monoclonal antibody to human αv integrins (intetumumab) in combination with docetaxel and prednisone for the first-line treatment of patients with metastatic castration-resistant prostate cancer. *Ann. Oncol.* **24**, 329–336 (2013).
33. Younes, A. et al. International Working Group consensus response evaluation criteria in lymphoma (RECIL 2017). *Ann. Oncol.* **28**, 1436–1447 (2017).
34. Cheson, B. D. et al. Recommendations for initial evaluation, staging, and response assessment of Hodgkin and non-Hodgkin lymphoma: the Lugano classification. *J. Clin. Oncol.* **32**, 3059–3067 (2014).
35. Kostakoglu, L. et al. Recil 2017 criteria demonstrated similar prognostic value and detected a comparable treatment difference between obinutuzumab- and rituximab-chemotherapy compared with Cheson 2007 and Lugano 2014 criteria in patients with previously untreated advanced-stage follicular lymphoma [abstract]. *Blood* **136** (Suppl. 1), 25–26 (2020).
36. Berzaczy, D. et al. RECIL versus Lugano for treatment response assessment in FDG-avid non-Hodgkin lymphomas: a head-to-head comparison in 54 patients. *Cancers* **12**, 9 (2019).
37. Wen, P. Y. et al. Updated response assessment criteria for high-grade gliomas: Response Assessment in Neuro-Oncology Working Group. *J. Clin. Oncol.* **28**, 1963–1972 (2010).
38. Choi, H. et al. Correlation of computed tomography and positron emission tomography in patients with metastatic gastrointestinal stromal tumor treated at a single institution with imatinib mesylate: proposal of new computed tomography response criteria. *J. Clin. Oncol.* **25**, 1753–1759 (2007).
39. Lencioni, R. & Llovet, J. M. Modified RECIST (mRECIST) assessment for hepatocellular carcinoma. *Semin. Liver Dis.* **30**, 52–60 (2010).
40. Lin, N. U. et al. Response assessment criteria for brain metastases: proposal from the RANO group. *Lancet Oncol.* **16**, e270–e278 (2015).
41. Wolchok, J. D. et al. Guidelines for the evaluation of immune therapy activity in solid tumors: immune-related response criteria. *Clin. Cancer Res.* **15**, 7412–7420 (2009).
42. Bohnsack, O., Hoos, A. & Ludajic, K. Adaptation of the immune related response criteria: irrecist [abstract 1070P]. *Ann. Oncol.* **25** (Suppl. 4), iv369 (2014).
43. Seymour, L. et al. iRECIST: guidelines for response criteria for use in trials testing immunotherapeutics. *Lancet Oncol.* **18**, e143–e152 (2017).
44. Tazdait, M. et al. Patterns of responses in metastatic NSCLC during PD-1 or PDL-1 inhibitor therapy: comparison of RECIST 1.1, irRECIST and iRECIST criteria. *Eur. J. Cancer* **88**, 38–47 (2018).
45. Park, H. J. et al. Comparison of RECIST 1.1 and iRECIST in patients treated with immune checkpoint inhibitors: a systematic review and meta-analysis. *Cancers* **13**, 120 (2021).
46. Jaffe, C. C. Measures of response: RECIST, WHO, and new alternatives. *J. Clin. Oncol.* **24**, 3245–3251 (2006).
47. Brody, T. *Clinical Trials* 2nd edn, 331–376 (Academic, 2016).
48. Rao, A. V. et al. Age-specific differences in oncogenic pathway dysregulation in patients with acute myeloid leukemia. *J. Clin. Oncol.* **27**, 5580–5586 (2009).
49. Schultz, K. R. et al. Improved early event-free survival with imatinib in Philadelphia chromosome-positive acute lymphoblastic leukemia: a Children's Oncology Group study. *J. Clin. Oncol.* **27**, 5175–5181 (2009).
50. Cortes, J. E. et al. Results of dasatinib therapy in patients with early chronic-phase chronic myeloid leukemia. *J. Clin. Oncol.* **28**, 398–404 (2010).
51. Marcucci, G. et al. High expression levels of the *ETS*-related gene, *ERG*, predict adverse outcome and improve molecular risk-based classification of cytogenetically normal acute myeloid leukemia: a Cancer and Leukemia Group B study. *J. Clin. Oncol.* **25**, 3337–3343 (2007).
52. Hughes, T. P. et al. Long-term prognostic significance of early molecular response to imatinib in newly diagnosed chronic myeloid leukemia: an analysis from the International Randomized Study of Interferon and STI571 (IRIS). *Blood* **116**, 3758–3765 (2010).
53. Yin, J. et al. Evaluation of event-free survival as a robust end point in untreated acute myeloid leukemia (Alliance A151614). *Blood Adv.* **3**, 1714–1721 (2019).
54. Bellera, C. A. et al. Protocol of the Definition for the Assessment of Time-to-event Endpoints in CANcer trials (DATECAN) project: formal consensus method for the development of guidelines for standardised time-to-event endpoints' definitions in cancer clinical trials. *Eur. J. Cancer* **49**, 769–781 (2013).
55. Tyagi, P. & Chu, E. Adjuvant irinotecan regimens in combination with infusional 5-fluorouracil/leucovorin fail to improve outcomes in surgically resected colorectal cancer. *Clin. Colorectal Cancer* **5**, 86–88 (2005).
56. Tolaney, S. M. et al. Updated Standardized Definitions for Efficacy End Points (STEEP) in adjuvant breast cancer clinical trials: STEEP version 2.0. *J. Clin. Oncol.* **39**, 2720–2731 (2021).
57. Hudis, C. et al. Proposal for standardized definitions for efficacy end points in adjuvant breast cancer trials: the STEEP system. *J. Clin. Oncol.* **25**, 2127–2132 (2007).
58. Saad, E. D. & Katz, A. Progression-free survival and time to progression as primary end points in advanced breast cancer: often used, sometimes loosely defined. *Ann. Oncol.* **20**, 460–464 (2009).
59. Prasad, V., Kim, C., Burotto, M. & Vandross, A. The strength of association between surrogate end points and survival in oncology: a systematic review of trial-level meta-analyses. *JAMA Intern. Med.* **175**, 1389–1398 (2015).
60. Ajani, J. A. et al. Disease-free survival as a surrogate endpoint for overall survival in adults with resectable esophageal or gastroesophageal junction cancer: a correlation meta-analysis. *Eur. J. Cancer* **170**, 119–130 (2022).
61. Google Scholar. Top publications: oncology. *Google Scholar* https://scholar.google.co.uk/citations?view_op=top_venues&hl=en&vq=med_oncology (2023).
62. Kip, K. E., Hollabaugh, K., Marroquin, O. C. & Williams, D. O. The problem with composite end points in cardiovascular studies: the story of major adverse cardiac events and percutaneous coronary intervention. *J. Am. Coll. Cardiol.* **51**, 701–707 (2008).
63. Anderson, R. L. et al. A framework for the development of effective anti-metastatic agents. *Nat. Rev. Clin. Oncol.* **16**, 185–204 (2019).
64. Beaver, J. A., Kluetz, P. G. & Pazdur, R. Metastasis-free survival – a new end point in prostate cancer trials. *N. Engl. J. Med.* **378**, 2458–2460 (2018).
65. Bellera, C. A. et al. Guidelines for time-to-event end point definitions in sarcomas and gastrointestinal stromal tumors (GIST) trials: results of the DATECAN initiative (Definition for the Assessment of Time-to-event Endpoints in CANcer trials). *Ann. Oncol.* **26**, 865–872 (2015).
66. Amabile, S. et al. Clinical significance of Distant Metastasis-Free Survival (DMFS) in melanoma: a narrative review from adjuvant clinical trials. *J. Clin. Med.* **10**, 5475 (2021).
67. Pan, H. et al. Comparison of survival outcomes among patients with breast cancer with distant vs ipsilateral supraclavicular lymph node metastases. *JAMA Netw. Open* **4**, e211809 (2021).
68. Zhan, H., Zhao, X., Lu, Z., Yao, Y. & Zhang, X. Correlation and survival analysis of distant metastasis site and prognosis in patients with hepatocellular carcinoma. *Front. Oncol.* **11**, 652768 (2021).
69. Twelves, C. et al. "New" metastases are associated with a poorer prognosis than growth of pre-existing metastases in patients with metastatic breast cancer treated with chemotherapy. *Breast Cancer Res.* **17**, 150 (2015).
70. Mori, R. et al. The mode of progressive disease affects the prognosis of patients with metastatic breast cancer. *World J. Surg. Oncol.* **16**, 169 (2018).

71. Suzuki, C. et al. The initial change in tumor size predicts response and survival in patients with metastatic colorectal cancer treated with combination chemotherapy. *Ann. Oncol.* **23**, 948–954 (2012).
72. Stein, A. et al. Survival prediction in everolimus-treated patients with metastatic renal cell carcinoma incorporating tumor burden response in the RECORD-1 trial. *Eur. Urol.* **64**, 994–1002 (2013).
73. Haslam, A., Hey, S. P., Gill, J. & Prasad, V. A systematic review of trial-level meta-analyses measuring the strength of association between surrogate end-points and overall survival in oncology. *Eur. J. Cancer* **106**, 196–211 (2019).
74. Walsh, E. M., Nunes, R., Wilkinson, M. J. & Santa-Maria, C. A. Extended endocrine therapy for early-stage breast cancer: how do we decide? *Curr. Oncol. Rep.* **22**, 123 (2020).
75. Algorashi, I., Goldvaser, H., Ribnikar, D., Cescon, D. W. & Amir, E. Evolution in sites of recurrence over time in breast cancer patients treated with adjuvant endocrine therapy. *Cancer Treat. Rev.* **70**, 138–143 (2018).
76. Wilson, B. E., Desnoyers, A., Al-Showbaki, L., Nadler, M. B. & Amir, E. A retrospective analysis of changes in distant and breast cancer related disease-free survival events in adjuvant breast cancer trials over time. *Sci. Rep.* **12**, 6352 (2022).
77. Robinson, A. G. et al. Patient perspectives of value of delayed disease progression on imaging (imaging PFS). A treatment trade-off experiment. *J. Cancer Policy* **30**, 100301 (2021).
78. Brundage, M. D. et al. Patients' attitudes and preferences toward delayed disease progression in the absence of improved survival. *J. Natl Cancer Inst.*, https://doi.org/10.1093/jnci/djad138 (2023).
79. Cordoba, G., Schwartz, L., Woloshin, S., Bae, H. & Gøtzsche, P. C. Definition, reporting, and interpretation of composite outcomes in clinical trials: systematic review. *BMJ* **341**, c3920 (2010).
80. European Medicines Agency. Appendix 1 to the guideline on the evaluation of anticancer medicinal products in man. *European Medicines Agency* https://www.ema.europa.eu/en/documents/scientific-guideline/appendix-1-guideline-evaluation-anticancer-medicinal-products-man-methodological-consideration-using_en.pdf (2012).
81. European Network for Health Technology Assessment. Personalised medicine and co-dependent technologies, with a special focus on issues of study design. *EUNETHTA* https://www.eunethta.eu/wp-content/uploads/2018/03/Personalized_Medicine_2016-03-07_reflection_paper_pm_2nd_draft.pdf (2015).
82. Sopik, V. & Narod, S. A. The relationship between tumour size, nodal status and distant metastases: on the origins of breast cancer. *Breast Cancer Res. Treat.* **170**, 647–656 (2018).
83. Gray, J. W. Evidence emerges for early metastasis and parallel evolution of primary and metastatic tumors. *Cancer Cell* **4**, 4–6 (2003).
84. Bryce, A. H. et al. Radiographic progression with nonrising PSA in metastatic castration-resistant prostate cancer: post hoc analysis of PREVAIL. *Prostate Cancer Prostatic Dis.* **20**, 221–227 (2017).
85. Merino, M. et al. Irreconcilable differences: the divorce between response rates, progression-free survival, and overall survival. *J. Clin. Oncol.* **41**, 2706–2712 (2023).
86. Kumar, S. K. et al. Venetoclax or placebo in combination with bortezomib and dexamethasone in patients with relapsed or refractory multiple myeloma (BELLINI): a randomised, double-blind, multicentre, phase 3 trial. *Lancet Oncol.* **21**, 1630–1642 (2020).
87. D'Alterio, C., Scala, S., Sozzi, G., Roz, L. & Bertolini, G. Paradoxical effects of chemotherapy on tumor relapse and metastasis promotion. *Semin. Cancer Biol.* **60**, 351–361 (2020).
88. Kim, H. T., Logan, B. & Weisdorf, D. J. Novel composite endpoints after allogeneic hematopoietic cell transplantation. *Transplant. Cell. Ther.* **27**, 650–657 (2021).
89. Soiffer, R. J. & Chen, Y.-B. Pharmacologic agents to prevent and treat relapse after allogeneic hematopoietic cell transplantation. *Blood Adv.* **1**, 2473–2482 (2017).
90. Oulhaj, A., El Ghouch, A. & Holman, R. R. Testing for qualitative heterogeneity: an application to composite endpoints in survival analysis. *Stat. Methods Med. Res.* **28**, 151–169 (2019).
91. Carroll, K. J. Analysis of progression-free survival in oncology trials: some common statistical issues. *Pharm. Stat.* **6**, 99–113 (2007).
92. Herbst, R. S. et al. Adjuvant osimertinib for resected EGFR-mutated stage IB-IIIA non-small-cell lung cancer: updated results from the phase III randomized ADAURA trial. *J. Clin. Oncol.* **41**, 1830–1840 (2023).
93. Chen, E. Y., Joshi, S. K., Tran, A. & Prasad, V. Estimation of study time reduction using surrogate end points rather than overall survival in oncology clinical trials. *JAMA Intern. Med.* **179**, 642 (2019).
94. Marsal, J.-R. et al. The use of a binary composite endpoint and sample size requirement: influence of endpoints overlap. *Am. J. Epidemiol.* **185**, 832–841 (2017).
95. Cameron, L. *Multiple Endpoints in Randomized Controlled Trials: a Review and an Illustration of the Global Test*. Thesis, Western (2023).
96. Faber, J. & Fonseca, L. M. How sample size influences research outcomes. *Dent. Press J. Orthod.* **19**, 27–29 (2014).
97. Bonnetain, F. et al. Guidelines for time-to-event end-point definitions in trials for pancreatic cancer. Results of the DATECAN initiative (Definition for the Assessment of Time-to-event End-points in CANcer trials). *Eur. J. Cancer* **50**, 2983–2993 (2014).
98. Kramar, A. et al. Guidelines for the definition of time-to-event end points in renal cell cancer clinical trials: results of the DATECAN project. *Ann. Oncol.* **26**, 2392–2398 (2015).
99. Reck, M. et al. Pembrolizumab versus chemotherapy for PD-L1-positive non-small-cell lung cancer. *N. Engl. J. Med.* **375**, 1823–1833 (2016).
100. Kantoff, P. W. et al. Hydrocortisone with or without mitoxantrone in men with hormone-refractory prostate cancer: results of the cancer and leukemia group B 9182 study [abstract]. *J. Clin. Oncol.* **17** (8), 2506 (1999).
101. Beer, T. M. et al. Enzalutamide in metastatic prostate cancer before chemotherapy. *N. Engl. J. Med.* **371**, 424–433 (2014).
102. Durie, B. G. M. et al. Bortezomib with lenalidomide and dexamethasone versus lenalidomide and dexamethasone alone in patients with newly diagnosed myeloma without intent for immediate autologous stem-cell transplant (SWOG S0777): a randomised, open-label, phase 3 trial. *Lancet* **389**, 519–527 (2017).
103. O'Brien, M. et al. Pembrolizumab versus placebo as adjuvant therapy for completely resected stage IB–IIIA non-small-cell lung cancer (PEARLS/KEYNOTE-091): an interim analysis of a randomised, triple-blind, phase 3 trial. *Lancet Oncol.* **23**, 1274–1286 (2022).
104. Van Cutsem, E. et al. Randomized phase III trial comparing biweekly infusional fluorouracil/leucovorin alone or with irinotecan in the adjuvant treatment of stage III colon cancer: PETACC-3. *J. Clin. Oncol.* **27**, 3117–3125 (2009).
105. Martin, M. et al. Neratinib after trastuzumab-based adjuvant therapy in HER2-positive breast cancer (ExteNET): 5-year analysis of a randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet Oncol.* **18**, 1688–1700 (2017).
106. Stone, R. M. et al. Midostaurin plus chemotherapy for acute myeloid leukemia with a FLT3 mutation. *N. Engl. J. Med.* **377**, 454–464 (2017).
107. Forde, P. M. et al. Neoadjuvant nivolumab plus chemotherapy in resectable lung cancer. *N. Engl. J. Med.* **386**, 1973–1985 (2022).
108. Shah, B. D. et al. KTE-X19 for relapsed or refractory adult B-cell acute lymphoblastic leukaemia: phase 2 results of the single-arm, open-label, multicentre ZUMA-3 study. *Lancet* **398**, 491–502 (2021).
109. Luke, J. J. et al. Pembrolizumab versus placebo as adjuvant therapy in completely resected stage IIB or IIC melanoma (KEYNOTE-716): a randomised, double-blind, phase 3 trial. *Lancet* **399**, 1718–1729 (2022).

Author contributions

All authors contributed substantially to discussion of the content. A.W. researched data for the article and wrote the manuscript. J.T. contributed to drafting of the figures. A.W. and V.P. reviewed and edited the manuscript before submission.

Competing interests

V.P. receives research funding from Arnold Ventures through a grant made to UCSF and royalties for books and writing from Johns Hopkins Press, MedPage and the Free Press; and declares consultancy roles with UnitedHealthcare and OptumRX. In addition, V.P. hosts the podcasts, Plenary Session, VPZD and Sensible Medicine; writes the newsletters, Sensible Medicine, the Drug Development Letter, and VP's Observations and Thoughts; and runs the YouTube channel Vinay Prasad MD MPH, which collectively earn revenue on the following platforms: Patreon, YouTube and Substack. The other authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41571-023-00823-5.

Peer review information *Nature Reviews Clinical Oncology* thanks M. Mayerhoefer, G. R. Mohyuddin and the other, anonymous, reviewers for their contribution to the peer review of this work.

[1]School of Medicine, University of California, San Francisco, San Francisco, CA, USA. [2]Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, CA, USA.