

<https://doi.org/10.1038/s41698-025-01231-x>

# Benchmarking progression-free survival ratio as primary endpoint in precision oncology clinical trials

Check for updates

Federico Nichetti<sup>1,2,20</sup> ✉, Jennifer Hüllelein<sup>2,20</sup>, Pauline du Rusquec<sup>3,4</sup>, Erin Pleasance<sup>5</sup>, Li Chen<sup>6,7</sup>, Andreas Mock<sup>8</sup>, Peter Horak<sup>9,10</sup>, Paolo Ambrosini<sup>11</sup>, Simone Rota<sup>11</sup>, Monica Niger<sup>11</sup>, Luca Agnelli<sup>12</sup>, Gabriele Tinè<sup>13</sup>, Dominic Edelmann<sup>14,15</sup>, Christophe Le Tourneau<sup>3,4,16</sup>, Janessa Laskin<sup>17</sup>, Giancarlo Pruneri<sup>11,18</sup>, Chi Wang<sup>6,7</sup>, Daniel Hübschmann<sup>2,9,19</sup>, Filippo de Braud<sup>1,18</sup>, Stefan Fröhling<sup>9,10,20</sup> & Luigi Mariani<sup>13,20</sup>

Progression Free Survival Ratio (PFSratio), as defined as the ratio between PFS on investigational treatment (PFS2) and PFS on the last prior therapy (PFS1), is a popular endpoint in precision oncology (PO) studies. In this work, five methodologies for PFSratio-based trial analysis (count-based, Kaplan Meier, Kernel-based Kaplan Meier, parametric and midrank) and two for trial design (GBVE and Weibull) are benchmarked. The Kernel-based Kaplan Meier analysis is most recommended, as it handles informative censoring and does not require PFS1/PFS2 distribution assumptions. Sample size and power calculation methods perform best when applied to settings with expected high PFS1/PFS2 correlation and median ratio. Analysis of five clinical trials (MOSCATO 01, WINTHER, MASTER, SHIVA and POG570) from >800 patients revealed an overall weak PFS1/PFS2 correlation (Kendall's  $\tau$  range 0.17-0.35), and an asymptotically unbiased median  $S_{\text{PFSratio}}(\delta=1.3) = 33\%$  by means of the Kernel-based analysis, while other methods considerably deviated in studies with censoring rate >10%. This methodology is implemented in the *PROPHETS* R package and Shiny app.

Over the last decades, the widespread use of next generation sequencing (NGS) technologies has led to the advent of precision oncology (PO), defined as the use of tumor molecular profiling to tailor anticancer treatment. Molecularly informed therapies have indeed often demonstrated better efficacy and tolerability than standard chemotherapy in clinical trials and in clinical practice<sup>1</sup>.

Typically, targeted therapies are initially tested in advanced treatment line settings when standard options are exhausted or have a low probability of success. Specifically, phase II trials represent the cornerstone to look for clinically significant activity of new compounds or combinations. However, most druggable molecular alterations are rare, with varying prevalence and actionability according to each tumor type. Moreover, in advanced

<sup>1</sup>Oncology Unit 1, Veneto Institute of Oncology IOV - IRCCS, Padua, Italy. <sup>2</sup>Computational Oncology, Molecular Precision Oncology Program, National Center for Tumor Diseases (NCT) and German Cancer Research Center (DKFZ), Heidelberg, Germany. <sup>3</sup>Department of Drug Development and Innovation (D3i), Institut Curie, Paris, France. <sup>4</sup>INSERM U900 Research unit, Institut Curie, Saint-Cloud, France. <sup>5</sup>Canada's Michael Smith Genome Sciences Centre, BC Cancer, Vancouver, BC, Canada. <sup>6</sup>Division of Cancer Biostatistics, Department of Internal Medicine, University of Kentucky, Lexington, KY, USA. <sup>7</sup>Markey Cancer Center, University of Kentucky, Lexington, KY, USA. <sup>8</sup>Institute of Pathology, University of Munich (LMU), Munich, Germany. <sup>9</sup>German Cancer Consortium (DKTK), Heidelberg, Germany. <sup>10</sup>Division of Translational Medical Oncology, National Center for Tumor Diseases (NCT) Heidelberg, German Cancer Research Center (DKFZ), Heidelberg, Germany. <sup>11</sup>Medical Oncology Department, Fondazione IRCCS Istituto Nazionale dei Tumori di Milano, Milan, Italy. <sup>12</sup>Department of Advanced Diagnostics, Fondazione IRCCS Istituto Nazionale dei Tumori di Milano, Milan, Italy. <sup>13</sup>Biostatistics for Clinical Research Unit, Epidemiology and Data Science Unit, Fondazione IRCCS Istituto Nazionale dei Tumori di Milano, Milan, Italy. <sup>14</sup>Division of Biostatistics, German Cancer Research Center (DKFZ), Heidelberg, Germany. <sup>15</sup>NCT Trial Center, National Center for Tumor Diseases, Heidelberg, Germany. <sup>16</sup>Paris-Saclay University, Paris, France. <sup>17</sup>Department of Medical Oncology, BC Cancer, Vancouver, BC, Canada. <sup>18</sup>Department of Oncology and Hemato-Oncology, University of Milan, Milan, Italy. <sup>19</sup>Pattern Recognition and Digital Medicine Group, Heidelberg Institute for Stem Cell Technology and Experimental Medicine (HI-STEM), Heidelberg, Germany. <sup>20</sup>These authors contributed equally: Federico Nichetti, Jennifer Hüllelein, Stefan Fröhling, Luigi Mariani. ✉e-mail: [federico.nichetti@iov.veneto.it](mailto:federico.nichetti@iov.veneto.it)

treatment lines, the use of placebo- or low effect standard of care-based control arms is ethically questionable. These features prevent from rapidly conducting numerically large phase II trials, especially randomized ones. Furthermore, in the lack of appropriate historical controls, single-arm phase II studies can be burdened by great interpatient and between-trial variability, so that the results are often difficult to generalize. These considerations are leading to a paradigm shift in PO phase II trials, with novel adaptive design and new endpoints being implemented in recent years<sup>2</sup>.

In this light, the Progression Free Survival Ratio (PFSratio) has been initially proposed by Von Hoff<sup>3</sup> as a novel endpoint for trials investigating new anticancer agents in patients who have failed at least one previous regimen. The PFSratio, also known as time to progression ratio (TTP<sub>r</sub>) or growth modulation index (GMI) is defined as the ratio between PFS on the therapy under investigation (PFS2) and the PFS on the last prior therapy (PFS1). This endpoint has some advantages compared to those usually adopted in phase II trials, like objective response rate or PFS, since it allows a) reduction in heterogeneity, as each patient serves as his/her own control, b) design of single-arm trials with limited sample size, and c) provision of a clinically relevant estimate of benefit of a new treatment. In recent years, a detailed exploration of PFSratio statistical properties has been performed, to correctly adopt it for the design and subsequent analysis of a trial<sup>4-7</sup>, but the methods remained scattered and difficult to interpret for clinicians. Moreover, several PO trials explored PFSratio as a measure of benefit, but different methods have been adopted, making the results poorly comparable.

In this work, we summarize current knowledge on PFSratio use in PO clinical trials, describing its clinical and statistical characteristics. Furthermore, we collect and implement different methods for the design and analysis of trials that use PFSratio as their primary endpoint. Finally, we explore PFSratio on outcome data of major PO clinical trials, providing a uniform analysis for proper interpretation of this methodology and corresponding results.

## Results

A worked example on how to properly design, conduct and analyze a clinical trial with PFSratio as its primary endpoint is provided in Fig. 1.

In clinical research, trial design with sample size calculation is the natural first step, while analysis of study results occurs only after its completion. In this work however, to guide the reader to the understanding of PFSratio methodology, we first describe the methods for PFSratio analysis and only later those for PFSratio-based trial design and sample size calculation.

The scenario considered here is that of a phase II trial that evaluates clinical benefit by comparing, within each treated patient, the failure time observed for a new agent under investigation with the failure time observed in the last prior line of treatment. The failure time is typically progression free survival (PFS), and the pair of values available for each patient is denoted as PFS1 and PFS2 for the two chronologically ordered treatment lines. Typically patients cannot be censored on PFS1, as only the patients who have progressed on the prior line of treatment will move to the subsequent line of treatment, but are censored on PFS2 whenever follow-up stops before progression for whatever reason, like the occurrence of censoring events (e.g. toxicity) or when the end of study has been reached. Note that, for PFSratio as an endpoint to be unbiased, the gold standard setting for its use is a clinical trial in which both PFS1 and PFS2 are assessed prospectively and in a consistent manner -that is, using the same criteria (radiological and/or clinical) and at the same evaluation frequency<sup>8</sup>.

### Exploration of paired failure times

Before PFSratio analysis, individual assessment of PFS1 and PFS2 is essential. In detail, it is of primary importance to assess the pattern of each PFS, the censoring rate of PFS2, and possibly the correlation between the two-paired PFS outcomes.

Descriptive analyses of PFS1 and PFS2 separately can be performed with conventional tools, like the Kaplan Meier method for plotting survival curves, while censoring rate is easily calculated as the proportion of patients

with censored PFS2 out of the total number. Estimating the correlation between PFS1 and PFS2 is a less easy task, as it involves the use of frailty models. In fact, positive correlation is likely to exist, as PFS1 and PFS2 may be affected not only by treatment efficacy, but also by other factors which are characteristics of each patient<sup>6</sup>. For example, patients with indolent tumors tend to stay longer under treatment, independently of therapy line, so that longer PFS1 corresponds to longer PFS2 or vice versa. However, this assumption does not always hold true<sup>9</sup>, especially in PO trials where a targeted agent may provide a significantly different benefit than the previous standard treatment, and must be therefore verified.

Regarding frailty models, they are extensions of conventional survival models that include additional “frailty” parameters that handle the correlation. One such model that is widely considered in this context is the Weibull Gamma frailty model, so named because the frailty term is distributed according to a gamma distribution, while, conditional on the frailty term, failure times PFS1 and PFS2 are assumed to follow a Weibull distribution. Interestingly, by denoting the frailty term with  $\beta\theta$ , this model allows estimating the inpatient dependence (i.e. the correlation between PFS1 and PFS2) via Kendall's  $\tau$ , with  $\tau = \theta/(\theta+2)$ . The calculation of this correlation index is crucial as it affects the performance of PFSratio in the analysis phase and sample size calculation in study design, as shown later.

The Weibull part of the model contains instead two parameters, one modeling the divergence between PFS1 and PFS2, the other one ( $K$ ) the shape of the hazard function, that is the rate at which the events occur over time. A value of this parameter  $>1$  ( $<1$ ) denotes a rate that is increasing (decreasing) over time, while a value equal to one implies an exponential distribution of failure times, characterized by a failure rate that is constant. These properties make the Weibull Gamma frailty model sufficiently flexible to fit most of the situations that we encounter in practical situations. A way to check the Weibull assumption is to plot the logarithm of the cumulative hazard, i.e.  $\log[-\log(S(t))]$  where  $\log$  stands for the natural logarithm and  $S(t)$  is the Kaplan Meier survival estimate for PFS, against  $\log(\text{survival time})$ . The Weibull assumption is correct if this plot gives two parallel and approximately straight lines for PFS1 and PFS2.

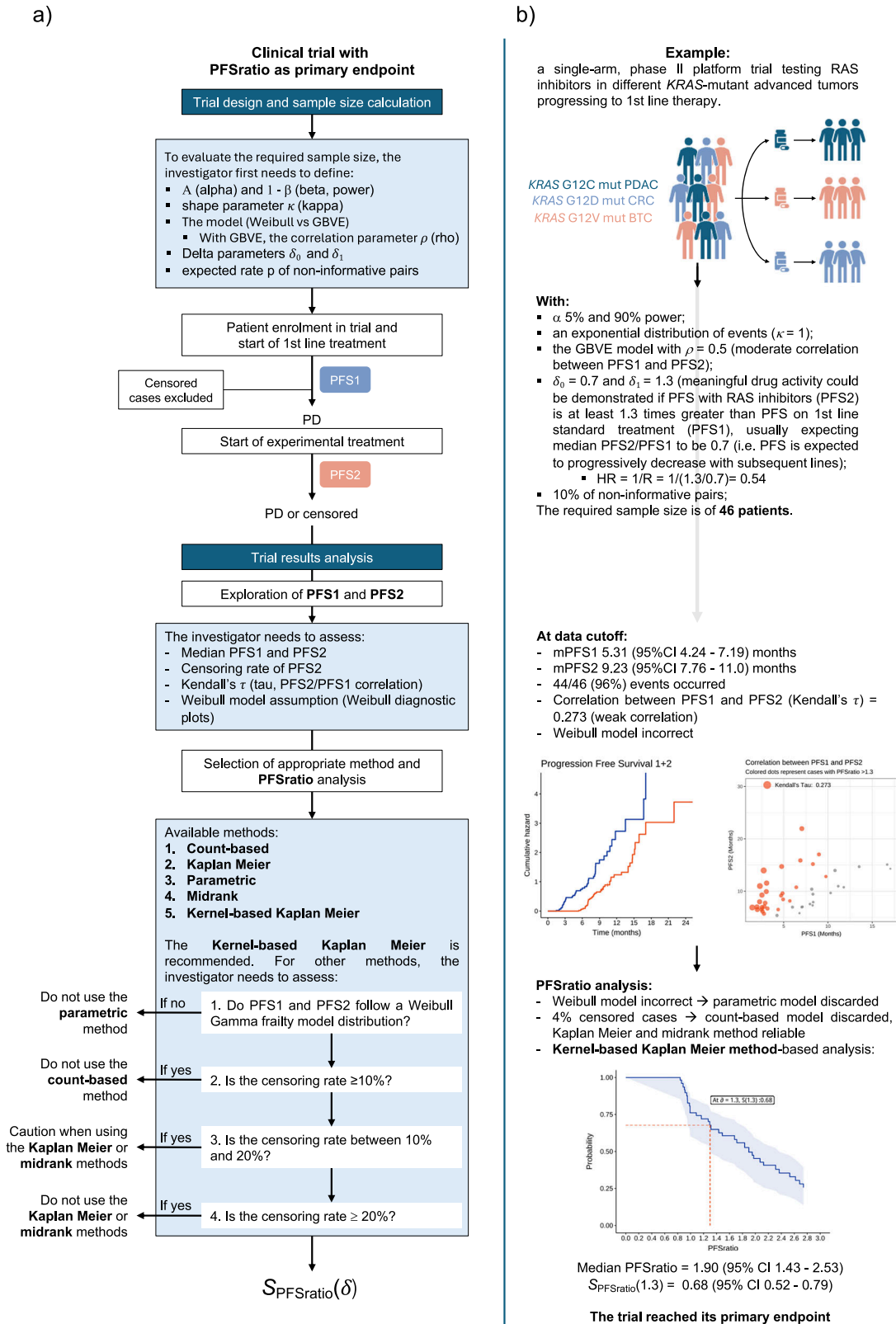
This exploration, along with considerations regarding censoring rates (discussed below), is necessary to apply the correct method for PFSratio analysis. In particular, among the methods described below, the parametric method can be properly applied only if PFS1 and PFS2 follow a Weibull distribution<sup>10</sup>.

### Analysis of PFSratio

As the name implies, PFSratio is calculated as PFS2/PFS1. In this way, a pair of values is transformed into a single but possibly censored, non-negative quantity, and a clinical study will provide  $N$  independent ratios computed from the pairs (PFS1, PFS2).

To judge the goodness of this ratio, a reference  $\delta$  reflecting minimal improvement in PFS with the investigated treatment needs to be set by the investigator. The  $\delta$  parameter represents the PFSratio cutoff that defines the investigational treatment as “effective” / the individual patient as “responder” if his/her PFSratio is no less than, or “ineffective” / “non responder” otherwise. A quite straightforward choice is  $\delta=1$ , because this value is exceeded whenever PFS2 is greater than PFS1, which is clearly a sign of therapeutic improvement. As more fully described below, however, other criteria for choosing  $\delta$  may be followed. Whatever the choice of  $\delta$ , a study that is based on PFSratio as an efficacy endpoint is aimed at estimating the probability that this ratio is equal to or greater than  $\delta$ , i.e.  $P(\text{PFSratio} \geq \delta)$ , or  $S_{\text{PFSratio}}(\delta)$  more compactly.  $S_{\text{PFSratio}}(\delta)$  (and its confidence interval, CI) thus represents the probability of having a ratio  $> \delta$ , which can be interpreted as the fraction of patients from a given cohort with a clinically relevant improvement of PFS2 relative to PFS1<sup>5,6</sup>.

To further determine the efficacy of the investigational treatment in a patient population, a targeted minimal proportion of patients for whom the treatment is effective, i.e., patients with PFSratio  $\geq \delta$ , needs to be set by the investigator. We refer to  $p_0$  as the targeted minimal proportion. A study is supportive of treatment efficacy whenever the probability of PFSratio  $\geq \delta$ ,



**Fig. 1 | PFSratio workflow. a, b** Proposed workflow (a) and worked example (b) to properly design, conduct and analyze a clinical trial with PFSratio as its primary endpoint. A single-arm, phase II platform trial testing novel RAS inhibitors in different KRAS-mutant advanced tumors progressing to 1st line therapy is hypothesized, and results are generated through a simulation with randomly generated data. The first step is trial design and sample size calculation. Then, the optimal setting is a to start enrolling patients at 1st line treatment initiation, following the same tumor assessment schedule during PFS1 and PFS2. Once the study has

adequate follow up, exploration of PFS1 and PFS2 is recommended, followed by PFSratio analysis selecting the proper method to avoid biased results. Abbreviations: BTC biliary tract cancer, CI confidence interval, CRC colorectal cancer, GBVE Gumbel's type B bivariate extreme-value, kappa shape of the hazard function, KM Kaplan Meier, PFS progression free survival, PFSratio progression free survival ratio, PD progressive disease, PDAC pancreatic ductal adenocarcinoma, rho pearson correlation index between PFS1 and PFS2, R median ratio between PFS1 and 2, detailed with corresponding δ<sub>1</sub> and δ<sub>0</sub>.

i.e.,  $S_{\text{PFSratio}}(\delta)$ , together with its lower CI, is shown to be greater than  $p_0$ . The choice of  $p_0$  varied across different studies. In MOSCATO 01<sup>11</sup>,  $p_0$  was set to be 15%. In WINTHER<sup>12</sup>,  $p_0$  was set to be 40% or 50%.

Several methods have been proposed to estimate  $S_{\text{PFSratio}}(\delta)$ , namely the count-based, Kaplan Meier, parametric, midrank (non-parametric) and a novel kernel-based Kaplan Meier, individually discussed below.

Note that, when selecting methods for PFSratio analysis, PFS2 censoring rates must be taken into account. We recommend using PFSratio as an endpoint when the censoring rate for PFS2 is below 50%, as estimate accuracy may otherwise decline -especially with higher  $\delta$  values and/or smaller median PFSratios. The kernel-based Kaplan-Meier method must be preferred when the censoring rate is  $\geq 20\%$ , as detailed below.

Most importantly, interpretation of PFSratio must also consider the clinical context, including the chosen  $\delta$  value and the cohort follow-up. Indeed, even with high censoring, censored cases with PFSratio  $> \delta$  remain informative.

### Count based method

The simplest estimator for  $S_{\text{PFSratio}}(\delta)$  is based on the count of PFSratio outcomes equal to or greater than  $\delta$ . For the proportion so obtained, a confidence interval and a statistical test can be constructed. A limitation of this method is that all those pairs for which PFSratio  $< \delta$  and PFS2 is censored are not informative. The reason is that actual PFS2, if known, could take values that make the ratio greater or less than  $\delta$ , thus generating uncertainty on how to handle these pairs. For this reason, they must be excluded from the calculation (unlike what happens with all other methods), with a loss of information as a consequence.

### Kaplan Meier method

The PFSratio can be regarded as the value of PFS2 measured in terms of PFS1 units: if PFS2 is censored, the PFSratio is censored too. For this reason, PFSratio can be treated as a time-to-event variable as the probability of surviving (or not progressing) at a given timepoint  $\delta$  and survival analysis methods can be applied for its evaluation. Among them, the Kaplan Meier method allows the investigator to easily obtain an estimate of  $S_{\text{PFSratio}}(\delta)$  with the corresponding CI. Furthermore, the Kaplan Meier survival curve describing the distribution of PFSratios in the overall study cohort can be plotted, and estimates of the median PFSratio and its corresponding CI can be straightforwardly obtained.

### Parametric method

It has been shown that if PFS1 and PFS2 follow the Weibull Gamma frailty model, their ratio follows a log-logistic distribution<sup>6,13</sup>. Using this distribution, the investigator can obtain maximum likelihood estimates of the distribution parameters from the possibly censored PFSratio data and derive  $S_{\text{PFSratio}}(\delta)$  by plugging these estimates into a simple formula.

### Midrank method

This method, as is typical of non-parametric tools, aims to compare the distributions of two independent variables without making strong statistical assumptions on their shape. In our specific context, the approach consists in using the ranks of each pair (PFS1, PFS2) to estimate  $S_{\text{PFSratio}}(\delta)$ . Due to censoring, the ranks of some observations are unknown but can be estimated by midranks according to the procedure proposed by Hudgens and Satten<sup>5,6,14</sup>. However, this midrank method requires the assumption that the censoring distributions for PFS1 and PFS2 are the same, which is violated because PFS1 is uncensored while PFS2 is censored<sup>14,15</sup>.

### Kernel-based Kaplan Meier method

Recently Chen et al.<sup>15</sup> recognized that, unlike the censoring of PFS2, which can usually be assumed to be uninformative, the censoring of PFSratio is always informative regardless of whether PFS2 is uninformatively censored or not. In detail, let  $C_2$  denote the time to the censoring events, such as toxicity or ending of study, corresponding to PFS2. As a result, the censoring variable corresponding to PFSratio, PFS2/PFS1, is  $C_2/\text{PFS1}$ . Since both

PFSratio and its censoring variable involve PFS1, they are dependent, which violates the uninformative (or independent) censoring assumption required by common survival analysis methods like those previously described. Consequently, novel parametric and non-parametric kernel-based Kaplan Meier methods were proposed for estimating the  $S_{\text{PFSratio}}(\delta)$  function and the corresponding confidence interval while adequately accounting for informative censoring. Unlike the Kaplan-Meier estimator, which directly estimate  $S_{\text{PFSratio}}(\delta)$ , i.e. the survival function of PFSratio, the kernel-based Kaplan-Meier estimator utilizes a two-stage procedure. In the first stage, it utilizes a kernel conditional Kaplan-Meier method to estimate the conditional survival function of the PFSratio given PFS1. In the second stage, it averages these conditional survival functions across all observed PFS1 values. Since the first stage requires only the assumption that the PFSratio is conditionally independently censored given PFS1, and the second stage imposes no additional assumptions, the consistency of the estimator relies solely on this conditional independent censoring assumption. Based on standard statistical theory, this assumption is satisfied if PFS2 is independently censored, a common and often reasonable condition in time-to-event analyses. Regarding the kernel function used in the estimator, Chen et al.<sup>9</sup> employed a modified Silverman Kernel, which is flatter than the Gaussian kernel and reduces the risk of near-zero values in the denominator, to ensure the numerical stability.

The authors proved theoretically and by means of extensive simulations that this new estimator is asymptotically unbiased. In contrast, previous methods were shown to be biased not only in simulations considering a censoring rate of 20% or 30%, but also in cases when the censoring rate was between 10% and 20%, especially for sample size  $\geq 100$  (data not reported).

Since the novel kernel-based non-parametric method holds the favorable features previously illustrated for the simple Kaplan Meier, while also accounting for dependent censoring, it will be used in the following sections as the reference method for trial analysis.

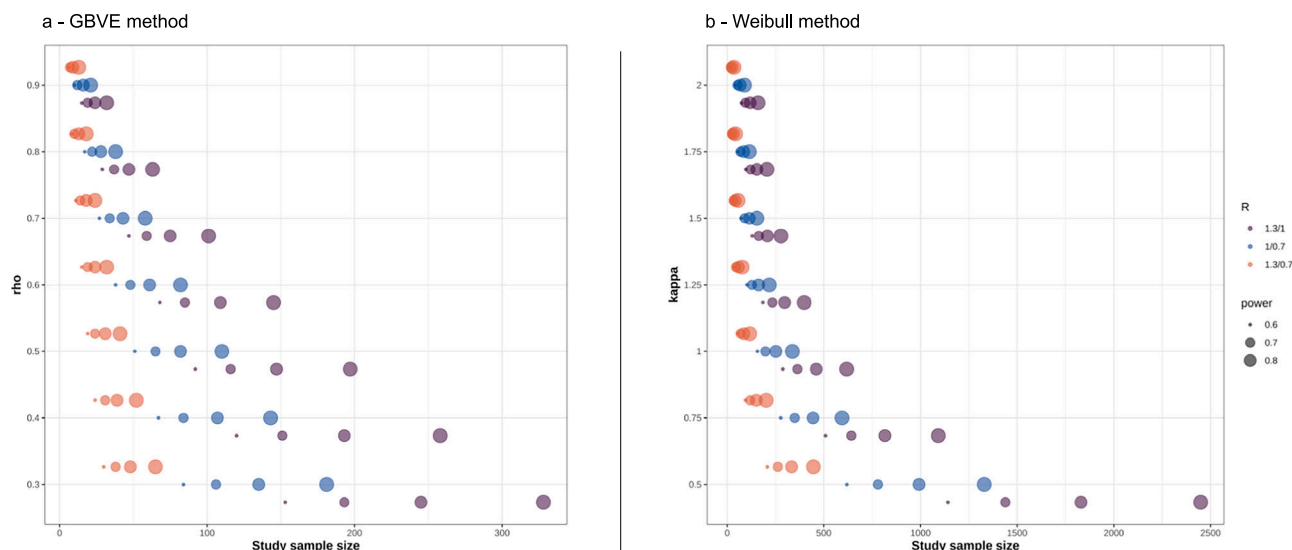
### Modified PFSratio

Concerning the duration of PFS1 and PFS2, a careful evaluation is recommended to avoid under- or over-estimation of PFSratio in cases of very short or long PFS1, respectively. To address this issue, Mock et al. proposed a modified PFSratio (mPFSratio), where PFS1 shorter than 2 months converts to 2 months and PFS2 longer than 6 months converts to 24 months before ratio calculation<sup>16</sup>. The underlying logic is that patients rapidly progressing with previous treatment, e.g. PFS1 of one month, might have an overly optimistic PFSratio even in case of short PFS2. Conversely, a sufficiently long PFS2, e.g. greater than six months, should in itself indicate sufficient activity of the experimental treatment. While the conversion of extremely short PFS1 to 2 months is reasonable in most cases, PFS2 conversion from  $>6$  to 24 months is debatable. In particular, 6 months can be taken as a clinically meaningful cutoff to consider a PFS2 treatment as effective, but it may change according to the clinical context and the chosen  $\delta$ . To address this issue, we propose a simple alternative method where PFS1 is equally transformed to 2 months in cases with extremely short ( $< 2$  months) PFS1, while PFS2 is transformed as follows:

- first, the clinical investigator decides what is the minimum PFS2 that can be considered satisfying irrespective of PFS1 (e.g., 6 months);
- then, once  $\delta$  is defined, in cases where PFS2 is satisfying but PFSratio is  $< \delta$ , PFS2 is transformed as  $\text{PFS1} * \delta + 0.25$ , with 0.25 representing a minimum advantage (~1 week, with survival estimates calculated in months) to PFS2 to avoid random fluctuations.

The adjustments to PFS1 and PFS2 reduce sensitivity to very small denominators and dampen outlier influence without imposing parametric structure. This leads to more consistent analysis results across the different estimators evaluated as compared to the results based on original PFSratio (see Supplementary Table 1).

Clearly, the use of the modified PFSratio must be even more strongly guided by the clinical context under consideration, in order to define



**Fig. 2 | PFSratio-based clinical trial design. a, b** Graphic representation of the impact of correlation ( $\rho$ ), power, HR on sample size in PFSratio-based clinical trial design. Abbreviations: GBVE Gumbel’s type B bivariate extreme-value, kappa shape

of the hazard function, rho pearson correlation index between PFS1 and PFS2, R median ratio between PFS1 and 2, detailed with corresponding  $\delta_1$  and  $\delta_0$ .

precisely what duration of PFS2 should be regarded as clinically meaningful, and to account for situations in which PFS1 may appear extremely short (e.g., clinical trials with CT scans scheduled every 6 weeks, which can result in very brief PFS1 estimates). Ultimately, since this is a descriptive, clinically derived endpoint rather than one built on a formal statistical model, further statistical validation will be required in this context.

**Sample size calculation**

Sample size calculation is a mandatory early step when designing a clinical trial, and PFSratio-based studies are no exception.

Once excluding simulation, which is generally possible but not easily doable, a handy solution is based on the use of analytical formulae, as proposed by Wu et al.<sup>7</sup> In practice, having defined the desired levels of significance ( $\alpha$ ) and power (1-beta), the calculation is based on the so-called Generalized Treatment Effect (GTE), that is the probability  $S_{PFSratio}(\delta)$  expected under the hypothesis of treatment efficacy. This probability can be specified directly, according to what the authors define as “Study design using generalized effect size”. More accurately, a GTE value can be derived on the basis of the assumptions regarding the distribution of PFS1 and PFS2.

One possibility is to assume the Weibull-Gamma frailty model previously described. With this model, specific values must be hypothesized for the shape parameter  $K$  and the ratio  $R$ , i.e., the median ratio of PFS2 versus PFS1. With  $K = 1$ , i.e., in case of exponential distribution,  $R$  corresponds to the reciprocal of the hazard ratio of PFS2 vs PFS1,  $HR = 1/R$ .

As an alternative, a bivariate exponential model, namely the Gumbel’s type B bivariate extreme-value (GBVE) model, may be assumed, in which both PFS1 and PFS2 follow an exponential distribution. In this case, the inputs required are the Pearson correlation coefficient  $\rho$  between PFS1 and 2 and the ratio  $R$ . Mick and colleagues<sup>4</sup> first showed that the  $\rho$  parameter is crucial: the stronger the correlation, the greater the power of the study and the smaller the sample size. Of note, a negative correlation ( $\rho < 0$ ) between PFS1 and 2 (i.e., settings in which a longer PFS2 corresponds to shorter PFS1 or vice versa) is clinically unlikely. Finally, regardless of the model assumed for sample size calculation, it is necessary to specify two values for  $\delta$ : one is the value assumed under the null hypothesis,  $\delta_0$ , which has the previously explained meaning of the cutoff that defines the starting point of effectiveness, the other is the value assumed under the alternative hypothesis,  $\delta_1$ , which represents the level of effectiveness at which the desired study power is achieved.

For example, PO trials aiming at demonstrating a meaningful benefit of targeted treatments after a prior standard therapy are usually designed with  $\delta_0 = 1$  and slightly greater  $\delta_1$  (e.g., 1.3 or 1.5), while in settings where PFS is expected to progressively decrease with subsequent lines of treatment, a PFS2 equal to PFS1 can be considered satisfactory, and therefore  $\delta_0$  can be set slightly below one (e.g., 0.7) and  $\delta_1 = 1$ .

Once the GTE has been specified or derived by either of the above two methods, the required number  $n$  of paired events can be calculated. To obtain the sample size  $N$ , i.e., the total number of pairs, adjustment of  $n$  by the expected rate  $p$  of non-informative pairs, that being the rate of cases with  $PFSratio < \delta$  and censored PFS2, based on the formula:  $N = n / (1 - p)$ , is required. If the trial is designed with a long enough follow-up time, the  $p$ -rate is small and the correction is minimal, usually around a few percentage points.

Figure 2a shows the impact of different  $\rho$ , power and  $R$  on the required  $N$  adopting the GBVE model: greater  $\rho$  values as well as greater  $R$  result in smaller patient numbers, while greater power requires larger numbers. From a clinical point of view, this exploration also shows that in settings with weak correlation (e.g.,  $< 0.3$ ), possibly combined with  $R$  close to one (small effect size), the required sample size becomes very large and makes it difficult to conduct the trial in a reasonable amount of time. Conversely, in contexts of expected clear superiority of PFS2 over PFS1 and with moderate/strong correlation between the two, a clinical study can be adequately conducted with only 20–30 patients. Regarding the Weibull model, Wu et al. previously showed that increasing  $K$  values result in smaller sample size<sup>7</sup>, while  $\rho$  is not directly considered in the calculation. For simulations presented in Fig. 2b, we show that higher  $K$ , power and  $R$  similarly result in smaller study sample size, while extremely high numbers (and thus not realistic for a PO trial) are required if  $K < 1$ .

An easy-to-use interface for the clinician to calculate the study sample size or power is available at <https://federico-nichetti.shinyapps.io/prophetsshiny/>. In addition, since many works explored PFSratio as a secondary analysis, investigators may be interested in testing the power of the analysis if  $N$  is known. To this aim, we implemented the reverse calculation, both with a GBVE and a Weibull model. Supplementary Fig. 1a, b show that, given a fixed sample size, studies with greater  $R$  (i.e.  $\delta_0 < 1$  and  $\delta_1 > 1$ ) keep a high power which is less affected by  $\rho$  and  $K$  in the GBVE and a Weibull models, respectively.

**Table 1 | Summary of PO clinical trials tested in the PFSratio-based analysis**

Study	year(s)	Site	Tumor types treated	Type of molecular screening	Primary endpoint	PFSratio analysis	n° of patients evaluable for PFSratio
<b>WINTHER</b> (NCT01856296)	2013–2015	Multicentric, international	Advanced cancers that progressed on standard Tx	Targeted genomic (236 cancer-related genes panel) + transcriptomic (microarrays)	PFSratio	Proportion of patients with PFSratio > 1.5	107
<b>MASTER</b> (NCT003127215)	2012–ongoing	Multicentric, national	Advanced rare cancers or cancers in young patients (< 51 years) that progressed on standard Tx	WGS/WES + transcriptomic (RNAseq)	Feasibility and clinical relevance of comprehensive molecular analysis in rare cancers	Proportion of patients with PFSratio > 1.3	255
<b>MOSCATO 01</b> (NCT01566019)	2011–2016	Monocentric	Advanced cancers that progressed on standard Tx	Targeted genomic (40–75 cancer-related genes panel) or WES + transcriptomic (RNAseq)	PFSratio	Proportion of patients with PFSratio > 1.3	194
<b>POG570</b> (NCT02155621)	2012–ongoing	Monocentric	Advanced cancers that progressed on standard Tx	WGS + transcriptomic (RNAseq)	Feasibility and clinical relevance of comprehensive molecular analysis in advanced cancers	Not done	190
<b>SHIVA</b> (NCT01771458)	2012–2014	Multicentric, national	Advanced cancers that progressed on standard Tx	Targeted genomic (45 cancer-related genes panel) + IHC for ER, PR and AR	PFS	Proportion of patients with PFSratio > 1.3	95

AR androgen receptor, ER estrogen receptor, IHC immunohistochemistry, PFSratio progression free survival ratio, PR progesterone receptor, Tx treatment, WES whole exome sequencing, WGS whole genome sequencing.

**PFSratio-based PO trials survey**

Data from five PO trials were collected and compared in the PFSratio-based analysis, namely MOSCATO 01<sup>11</sup>, WINTHER<sup>12</sup>, MASTER<sup>17</sup>, SHIVA<sup>18</sup>, and POG570<sup>19</sup>. All of these studies (except for POG570) previously adopted PFSratio as an endpoint, primary or secondary, and all shared the common feature of genome- and transcriptome-based treatment selection in advanced cancer patients. A summary of these studies is provided in Table 1.

**MOSCATO 01** (Molecular Screening for Cancer Treatment Optimization) was the first, monocentric, PO trial using targeted and whole exome (WES) plus transcriptome profiling to tailor treatment and adopting PFSratio as its primary endpoint<sup>11</sup>; **WINTHER** was an international, non-comparative trial in which genomic (via a 236 cancer-related genes panel, arm A) or transcriptomic (arm B) data were used to propose targeted therapies, which failed to meet its primary endpoint of PFSratio > 1.5 in 50% and 40% of arm A and B patients, respectively<sup>12</sup>. **MASTER** (Molecularly Aided Stratification for Tumor Eradication Research) is a multicentric registry trial and analytical platform established in 2012 by the German Cancer Consortium (DKTK) for prospective, multi-omics-guided stratification of patients with advanced cancers diagnosed at a young age (< 51 years) or with rare cancers, and comprises whole-genome/whole-exome sequencing (WGS/WES) and transcriptome profiling<sup>20</sup>. While the study is ongoing, the data presented here refer to the first published cohort treated up to November 2018 with available PFSratio data<sup>17</sup>. **SHIVA** was a multicentric, randomized phase II trial comparing treatment of physician choice (TPC) to molecularly-targeted agents (MTA, i.e., drugs that were approved for clinical use in France, but outside their indication) in patients with refractory cancers, allowing for crossover in both study arms<sup>21</sup>; here we applied PFSratio methods in the 95 patients who crossed-over, as previously reported<sup>18</sup>. Data analysis was performed separately on the TPC -> MTA and in the MTA -> TPC cohorts. **Personalized OncoGenomics (POG)** is a clinical trial conducted at BC Cancer (Canada) investigating the feasibility and utility of WGS and transcriptome analysis for informing targeted treatment of patients with advanced tumors. For this work, we analyzed for the first time the PFSratio of patients with available PFS1 and PFS2 data within the first 570 treated cases (POG570)<sup>19</sup>, with only the first targeted treatment considered in patients that received more than one molecularly guided therapy.

For this work, MOSCATO 01 PFS data were reconstructed from original paper plots using the *digitize* R package and results were double checked with data reconstructed using the *Inkscape* software; WINTHER data were retrieved, after formal request, from supplementary data of the original work<sup>12</sup> using the *pdftables* R package; MASTER data were shared after formal project approval at the DKTK MASTER Scientific Board meeting. SHIVA and POG570 data were shared after formal request to the corresponding authors of the original paper (SHIVA: P.d.R. and C.L.T., POG570: E.P. and J.L.)

Among the explored studies, a total of 841 patients were included, with as many pairs of PFS1 and 2. Details concerning each study are reported in Table 1. All studies included pretreated patients with advanced cancers, with varying sequencing coverage for molecular characterization, ranging from small targeted genomic panels to WGS and transcriptome sequencing. As shown, with the exception of POG570, these studies analyzed PFSratio simply as the proportion of cases with PFSratio > 1.3 or > 1.5.

We re-analyzed these trials according to the above-described methods. A summary of the analysis is provided in Table 2.

First, individual assessment of paired PFS1 and PFS2 was performed. Median PFS1 ranged from 2.0 to 4.1 months, while PFS2 from 2.0 to 3.5 months, with a censoring rate on PFS2 ranging from 1% to 20% (see Supplementary Fig. 2a–f).

Of note, Weibull diagnostic plots revealed that none of the studies can be considered as adequate for a parametric analysis, as shown in Supplementary Fig. 3a–f.

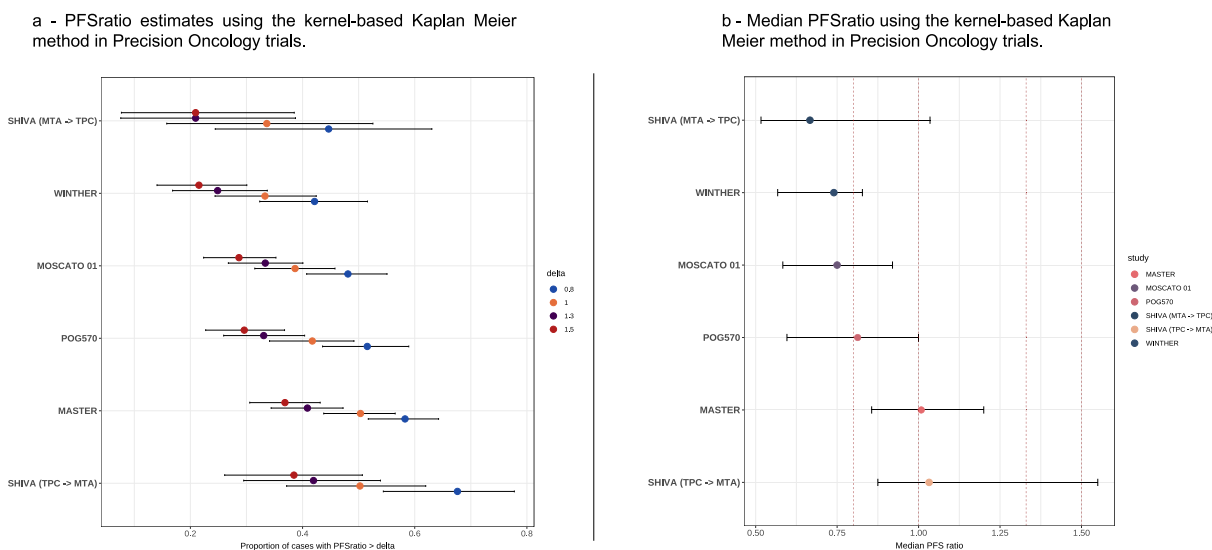
Moreover, analysis of PFS1 and PFS2 dependence showed that Kendall’s  $\tau$  ranged from 0.17 to 0.35, highlighting an overall positive though weak correlation (see Supplementary Fig. 4a–f).

**Table 2 | Summary of PFS1, PFS2 and PFSratio analysis of precision oncology trials**

Study	Median PFS1 (months, 95% CI)	Median PFS2 (months, 95% CI)	Median PFSratio	Weibull distribution	Kendall's Tau	Proportion of informative cases	$S_{PFSratio}(\delta=1.3)$
WINTHER	4.1 (3.2–5.2)	2.0 (1.8–3.1)	0.74 (0.57–0.83)	no	0.314	106/107 (99%)	0.25 (0.17–0.34)
MASTER	3.2 (3.0–3.7)	3.5 (3.1–4.1)	1.02 (0.86–1.21)	no	0.220	233/255 (91%)	0.41 (0.34–0.47)
MOSCATO 01	3.0 (2.7–3.6)	2.4 (1.9–2.8)	0.75 (0.58–0.92)	no	0.174	193/194 (99%)	0.33 (0.27–0.40)
POG570	3.5 (2.8–4.1)	3.4 (2.5–4.3)	0.81 (0.60–1.01)	no	0.228	152/190 (80%)	0.33 (0.26–0.41)
SHIVA (TPC -> MTA)	2.0 (1.9–4)	2.3 (2.1–4.1)	1.03 (0.87–1.55)	no	0.295	61/70 (87%)	0.42 (0.29–0.54)
SHIVA (MTA -> TPC)	2.8 (2.1–4.0)	2.3 (1.8–3.2)	0.67 (0.52–1.04)	no	0.354	24/25 (96%)	0.21 (0.08–0.39)

Median PFSratio and  $S_{PFSratio}(\delta=1.3)$  were calculated using the kernel-based Kaplan Meier method.

CI confidence interval, MTA molecularly targeted agent, PFS progression free survival, PFSratio progression free survival ratio, TPC treatment of physician's choice,  $S_{PFSratio}$  the probability that PFSratio is equal to or greater than  $\delta$ .



**Fig. 3 | PFSratio-based comparison of precision oncology clinical trials. a**  $S_{PFSratio}(\delta)$ , **b** median PFSratio Vertical dotted lines identify 0.8, 1, 1.3, and 1.5 points on the x axis. Abbreviations: MTA molecularly targeted agent, PFS progression free survival, PFSratio progression free survival ratio, TPC treatment of physician's choice.

PFSratio analysis was then performed with all five methods described above. First, we explored whether  $S_{PFSratio}(\delta)$  estimation was consistent between the methods with different  $\delta$  values (ranging from 0.1 to 3.0). Supplementary Fig. 5a–f plotted the  $S_{PFSratio}(\delta)$  curve based on different methods and Supplementary Table 1 reported the values of  $S_{PFSratio}(\delta)$  at  $\delta=1.3, 1.5$  and  $2.0$ .

The kernel-based Kaplan Meier method, being asymptotically unbiased, served as a reference for comparison. The curve based on the parametric method deviated considerably from that based on the kernel-based Kaplan Meier method in most studies, likely due to the violation of the parametric assumption. For the count-based, Kaplan Meier and midrank methods, which are all non-parametric, curves were very close to that based on the kernel-based Kaplan Meier method in studies with censoring rate < 10%, including WINTHER, MASTER, MOSCATO 01, and SHIVA (MTA -> TPC). However, differences became non-negligible in studies with censoring rate > 10%, including POG570 (20% censoring) and SHIVA (TPC -> MTA) (13% censoring). The count-based method substantially underestimated  $S_{PFSratio}(\delta)$  for  $\delta > 1.3$  in the SHIVA (TPC -> MTA) study. The midrank method overestimated  $S_{PFSratio}(\delta)$  in the SHIVA (TPC -> MTA) study, and substantially overestimated  $S_{PFSratio}(\delta)$  in the POG570 study. The Kaplan Meier method overestimated  $S_{PFSratio}(\delta)$  for  $\delta > 1$  in the POG570 study. A visual comparison of trials'  $S_{PFSratio}(\delta)$  and median PFSratio, as evaluated with the kernel-based Kaplan Meier method, is shown in Fig. 3a, b and Supplementary Fig. 6a–f.

The plots highlight that MASTER and the TPC -> MTA arm of SHIVA were associated with the best  $S_{PFSratio}(\delta)$  and median PFSratio.

We also explored the impact of adopting the modified PFSratio, for  $\delta=1.3$  and a minimum satisfying PFS2 of 6 months. Supplementary Table 1 shows that this method provides a slightly higher  $S_{PFSratio}(1.3)$  for WINTHER, MASTER and POG570 and a very similar or even smaller estimate for MOSCATO 01 and SHIVA, which is due to the high proportion of cases with PFS1 < 2 months in these trials.

Finally, given the weak correlation between PFS1 and PFS2 that we observed in the five studies, we performed a GBVE-based exploratory analysis aimed at evaluating the effective power of the PFSratio-based analysis and the sample size that would have been required to obtain a power of 90%. To this aim, we set  $\delta_1$  at 1.3 and 1.5, while  $\delta_0$  at 0.8 and 1, the expected proportion of non-informative pairs at 10% and we assumed Kendall's  $\tau$  as the  $\rho$  index of power/sample size calculation formula. As shown in Supplementary Table 2, most trials were adequately sized for a PFSratio-based analysis with  $\delta_1$  at 1.3 and  $\delta_0$  at 0.8, which represents the most rational clinical scenario. However, if  $\delta_0$  is set at 1 (or higher), with the observed PFS1-PFS2 correlations, the PFSratio-based analysis would be clearly underpowered, and much larger sample sizes would be required.

### Discussion

The main advantage of using PFSratio as a primary endpoint in clinical trials is the possibility of designing studies with good homogeneity (as each

patient acts as a control of him/herself) and with a limited sample size. This is indeed useful in PO studies, such as basket, umbrella or N-of-1 studies<sup>2,22</sup> in which each entity (patient, tumor type, target molecular alteration and targeted treatment) is rare and would therefore pose an accrual and time challenge in case of large, randomized studies. In the last 10 years, several molecularly driven clinical trials have demonstrated the potential to increase the therapeutic possibilities for patients with advanced cancers, with many adopting PFSratio as primary or secondary endpoint<sup>23</sup>. However, most trials did not adopt proper methods for study design and analysis of paired survival times, in part because statistical properties of this endpoint have not been widely explored. For example, among studies considered here, both MOSCATO 01 and WINTHER had PFSratio as primary endpoint, but by testing the  $S_{PFSratio}(\delta)$  against an arbitrary proportion, the authors concluded insufficient activity.

In this work, we summarized methods for PFSratio-based clinical trials design and analysis, so that these could be adopted in future studies. With regards to the analytical phase, among the five methods described, we consider the kernel-based Kaplan Meier as the most recommended, especially in contexts where the percentage of censored cases is high ( $> 20\%$ ), as it is intuitive for clinicians, able to provide an estimate of the median PFSratio and to account for dependent censoring. In the kernel-based Kaplan-Meier estimator, the use of a two-stage procedure combined with a kernel method strengthens its robustness and validity, as i) it explicitly handles the inherent dependent censoring of the PFSratio by conditioning on PFS1, enabling unbiased, fully nonparametric estimation of  $S_{PFSratio}(\delta)$ , ii) with a modified Silverman kernel, it ensures numerical stability under right-skewed PFS1 and iii) it results in systematically better calibrated curves than KM/naïve parametric estimates, with differences amplified with increasing censoring. Contrarily, despite its immediacy, the count-based method does not consider PFSratio as a time-to-event variable and thus ignores the impact of censored pairs. As such, it requires exclusion of cases with  $PFSratio \leq \delta$  if censored, while cases with  $PFSratio > \delta$  are kept as successful irrespective of the censoring status. Therefore, we consider this method as the least attractive, and it should not be adopted in settings with high censoring rates, as it may lead to incorrect exclusion of a significant number of patients from the analysis. The parametric method is limited by the assumption that PFS1 and PFS2 data need to follow a Weibull distribution, which is not always fulfilled. Nonetheless, we showed how overall results are very similar with the different methods in the five studies explored.

Concerning trial design, we implemented a user-friendly tool to calculate the required sample size (or the study power, if the sample is given) directly providing an expected GTE or evaluating it with two different methods (GBVE and Weibull). Among these, the GBVE method is influenced by the expected correlation between PFS1 and PFS2, but grants the design of PFSratio-based trial with adequate power and limited sample size in most cases. In our opinion, in a PO setting, a  $\delta_1 > 1$  (1.3 or 1.5, i.e., the molecularly informed treatment is expected to perform better than patient's last prior therapy) and  $\delta_0 < 1$  (i.e. if not effective, the molecularly informed treatment is expected to perform worse than patient's last prior therapy, as PFS is expected to deteriorate with subsequent lines of therapy) represents the optimal design to conduct a clinical trial on a N-of-1 population.

Given these tools, we performed a PFSratio-based survey of five PO clinical trials. First, our analysis highlighted a relevantly weak PFS1/PFS2 correlation in all studies, which might be due to the heterogeneity of tumor types and corresponding treatments included in each study. In this light, future studies are needed to elucidate which cancer types are associated with strongest intra-patient correlation and can thus be studied with small sample-sized, PFSratio-based clinical trials. Until such data become available, PO trials should be designed assuming a weak to moderate  $\rho$ .

Furthermore, we compared PFSratio-based results between trials as assessed by the kernel-based Kaplan Meier method. This analysis revealed POG570, MASTER and the TPC -> MTA SHIVA arm as studies with higher  $S_{PFSratio}(\delta)$  and median PFSratio. While a proper comparison between these trials cannot be done, we can hypothesize

some reasons for this difference: first, while these trials were overall conducted in a similar time frame, MOSCATO 01 and WINTHER results were published much earlier, so that targeted therapies options in these trials may have been limited compared to more recent studies. More likely, POG570 and MASTER patients were treated according to results of WGS/WES and transcriptome sequencing, as compared to targeted sequencing panels of the other studies, which may have provided more informative results for targeted therapies. Also, MASTER selectively includes patients with rare cancers or cancers in young patients, where molecularly informed therapies can have more impact as standard therapeutic options are even more limited. Finally, a separate consideration must be done for SHIVA, given the study design (see below) and since MTAs associated with longer PFSratio were represented by hormone therapies (tamoxifen or abiraterone) or everolimus, thus only partially comparing to other studies<sup>18</sup>.

PFSratio has limitations as primary endpoint for clinical trials. First, it must be noted that PFS2 is calculated from the first day of treatment administration, thus not considering the time that may elapse from the end of the prior treatment and the impact this may have on patients' status. Moreover, PFSratio has been mostly adopted in PO trials performed in patients progressing after exhaustion of standard treatment options, which may have provided the greatest OS benefit. Based on this, the PFSratio cannot be adopted as a granted surrogate endpoint for OS, and its exploration in this regard has been encouraging but limited<sup>17</sup> so far.

Most importantly, a reliable estimation of PFSratio is based on uniform disease assessment during the two lines of treatment considered, both in terms of frequency (e.g., CT scans regularly performed every 2 months)<sup>8</sup> and of techniques (radiological, clinical or laboratory assessment methods), as suggested in Fig. 1. This is questionable for both PFS1 and PFS2 in retrospective studies: in a recent meta-analysis on outcomes of treatments recommended by molecular tumor boards, Gladstone et al. highlighted the high variation in the required period to fulfill the criteria for defining stable disease, ranging from 6 weeks to 6 months, among the studies considered<sup>23</sup>, potentially resulting in poor interpretability of PFSratio. Similarly, clinical trials adopting PFSratio have frequently considered PFS1 retrospectively, which can result in estimation bias. For example, CT scans during prior standard treatment may have been performed at wider intervals, resulting in artificial lengthening of PFS1. Conversely, PFS1 may be assessed in the context of clinical trials with shorter evaluation intervals (e.g., CT scans every 6 weeks) or defined using non-radiological criteria (e.g., increases in tumor blood biomarkers), which can lead to PFSratio being distorted in the opposite direction. In this light, the cohort of crossing-over patients from the SHIVA trial represented the optimal setting for PFSratio exploration, as both PFS1 and PFS2 were assessed within the study, so that the striking difference in median PFSratio between the two SHIVA cohorts (TPC -> MTA vs MTA -> TPC) is not surprising. Therefore, we consider it mandatory that studies reporting a PFSratio-based analysis provide detailed information on the assessment methods and timing of evaluations for both PFS1 and PFS2.

Most importantly, statistical considerations must be balanced with clinical judgment to determine the appropriateness of PFSratio use in a given context. Although random variability may arise from the weak correlation between PFS1 and PFS2, PO provides an ideal setting for the adoption of PFSratio as an endpoint. In this context, targeted therapies (captured by PFS2) are specifically expected to surpass the efficacy of preceding non-targeted treatments (PFS1). This reversal -contrary to the physiological progressive shortening of PFS with successive lines of therapy - is precisely what indicates benefit of the targeted intervention. In this light, although not an explicit requirement, the interpretability of PFSratio in PO studies implicitly relies on the similarity of non-targeted treatments used for PFS1. Also, while very short or long PFS1 may reduce the clinical relevance of PFSratio, we showed that the use of a modified PFSratio results in estimate distortions, and should thus be carefully guided by the clinical setting and according to the specific study scenario.

Finally, so far, PFSratio has been used primarily as an endpoint in single-arm studies. However, future directions include the development of suitable methodologies to compare two or more treatment arms, incorporating multivariable analyses that adjust for relevant clinical and biological patient characteristics.

In conclusion, we provided a clinician-oriented overview of PFSratio statistical properties, which can represent a benchmark for future study design and analysis. The methods presented are meant for a phase II clinical trial, especially in the PO, N-of-1 setting where each patient represents a rare entity treated with molecularly informed therapies. We applied these methods to survey five PO clinical trials, highlighting differences and pitfalls, and implemented the methods in a user-friendly web application. With these tools, prospective trials are awaited to confirm the validity of PFSratio as a clinically meaningful endpoint in PO.

## Methods

The manuscript is meant as a guide for a reader with a clinical background in medical oncology and therefore avoid an in-depth statistical description with mathematical formulae, for which reference is made to dedicated works<sup>4–7,15,16</sup>. Stemming from these works, we wrapped and implemented the methods here presented in the *PROPHETS* R package and Shiny app ([https://github.com/fedenichetti/prophets\\_package](https://github.com/fedenichetti/prophets_package)) as a user-friendly web calculator for clinical purposes. All analyses later shown were performed in R (Version 4.2.0, 2022-04-22) using the RStudio software [Posit open source data science company].

## Data availability

Paired survival outcomes of the WINTHER<sup>12</sup> and MOSCATO 01<sup>11</sup> study are publicly available from original publications. For SHIVA<sup>21</sup>, MASTER<sup>17</sup> and POG570<sup>19</sup>, data are available upon reasonable request shared to the corresponding author and to corresponding authors of the original publications.

## Code availability

The code for the *PROPHETS* R package and Shiny app are available at [https://github.com/fedenichetti/prophets\\_package](https://github.com/fedenichetti/prophets_package).

Received: 14 May 2025; Accepted: 27 November 2025;

Published online: 15 December 2025

## References

- Subbiah, V. & Kurzrock, R. Challenging standard-of-care paradigms in the precision oncology era. *Trends Cancer* **4**, 101–109 (2018).
- Fountzilas, E., Tsimberidou, A. M., Vo, H. H. & Kurzrock, R. Clinical trial design in the era of precision medicine. *Genome Med.* **14**, 101 (2022).
- Von Hoff, D. D. There are no bad anticancer agents, only bad clinical trial designs—twenty-first Richard and Hinda Rosenthal Foundation Award Lecture. *Clin. Cancer Res.* **4**, 1079–1086 (1998).
- Mick, R., Crowley, J. J. & Carroll, R. J. Phase II clinical trial design for noncytotoxic anticancer agents for which time to disease progression is the primary endpoint. *Control Clin. Trials* **21**, 343–359 (2000).
- Kovalchik, S. & Mietlowski, W. Statistical methods for a phase II oncology trial with a growth modulation index (GMI) endpoint. *Contemp. Clin. Trials* **32**, 99–107 (2011).
- Texier, M. et al. Evaluation of treatment effect with paired failure times in a single-arm phase II trial in oncology. *Comput Math. Methods Med* **2018**, 1672176 (2018).
- Wu, J. et al. Phase II trial design with growth modulation index as the primary endpoint. *Pharm. Stat.* **18**, 212–222 (2019).
- Samaille, T. et al. Impact of the timing of tumor assessments on median progression-free survival in clinical trials in advanced cancer patients. *ESMO Open* **7**, 100366 (2022).
- Buyse, M. et al. Progression-free survival ratio as end point for phase II trials in advanced solid tumors. *J. Clin. Oncol.* **29**, e451–452 (2011).

- Zhang, Z. Parametric regression model for survival data: Weibull regression model as an example. *Ann. Transl. Med.* **4**, 484 (2016).
- Massard, C. et al. High-throughput genomics and clinical outcome in hard-to-treat advanced cancers: results of the Moscato 01 trial. *Cancer Discov.* **7**, 586–595 (2017).
- Rodon, J. et al. Genomic and transcriptomic profiling expands precision cancer medicine: the Winther trial. *Nat. Med.* **25**, 751–758 (2019).
- Owen, W. J. A power analysis of tests for paired lifetime data. *Lifetime Data Anal.* **11**, 233–243 (2005).
- Hudgens, M. & Satten, G. Midrank unification of rank tests for exact, tied, and censored data. *J. Nonparametr. Stat.* **14**, 569–581 (2002).
- Chen, L., Burkard, M., Wu, J., Kolesar, J. M. & Wang, C. Estimating the distribution of ratio of paired event times in phase II oncology trials. *Stat. Med.* **42**, 388–406 (2023).
- Mock, A. et al. Community-driven development of a modified progression-free survival ratio for precision oncology. *ESMO Open* **4**, e000583 (2019).
- Horak, P. et al. Comprehensive genomic and transcriptomic analysis for guiding therapeutic decisions in patients with rare cancers. *Cancer Discov.* **11**, 2780–2795 (2021).
- Belin, L. et al. Randomized phase II trial comparing molecularly targeted therapy based on tumor molecular profiling versus conventional therapy in patients with refractory cancer: cross-over analysis from the SHIVA trial. *Ann. Oncol.* **28**, 590–596 (2017).
- Pleasant, E. et al. Whole-genome and transcriptome analysis enhances precision cancer treatment options. *Ann. Oncol.* **33**, 939–949 (2022).
- Lier, A. et al. Validating comprehensive next-generation sequencing results for precision oncology: the NCT/DKTK molecularly aided stratification for tumor eradication research experience. *JCO Precis Oncol.* **2**, 1–13 (2018).
- Le Tourneau, C. et al. Molecularly targeted therapy based on tumour molecular profiling versus conventional therapy for advanced cancer (SHIVA): a multicentre, open-label, proof-of-concept, randomised, controlled phase 2 trial. *Lancet Oncol.* **16**, 1324–1334 (2015).
- Gouda, M. A., Buschhorn, L., Schneeweiss, A., Wahida, A. & Subbiah, V. N-of-1 trials in cancer drug development. *Cancer Discov.* OF1-OF9 (2023) <https://doi.org/10.1158/2159-8290.CD-22-1377>.
- Gladstone, B. P. et al. Systematic review and meta-analysis of molecular tumor board data on clinical effectiveness and evaluation gaps. *NPJ Precis Oncol.* **9**, 96 (2025).

## Acknowledgements

This work was supported by the NCT Molecular PO Program (Germany). Drs. Chen and Wang's work was supported by the National Cancer Institute (R21CA284179-01A1), USA.

## Author contributions

The work reported in the paper has been performed by the authors, unless clearly specified in the text. - Conceptualization: N.F. and M.L. Methodology: N.F., H.J., C.L., W.C., and M.L. Investigation: N.F. and M.L. Resources: all authors - Data curation: N.F., H.J., M.A., R.S., and A.P. Formal analysis: N.F., H.J., and M.L. Visualization: H.J. and N.F. Writing - original draft: N.F., M.L., and E.D. Writing - review & editing: all authors - Supervision: d.B.F., P.G., F.S., and H.D.

## Funding

Open Access funding enabled and organized by Projekt DEAL.

## Competing interests

Conflict of interest: F.N. personal fees from AstraZeneca, Accademia della Medicina, DOC CONGRESS SRL; travel grants from Merck.S.F.: Consulting or advisory board membership: Bayer, Illumina, Roche; honoraria: Amgen, Eli Lilly, PharmaMar, Roche; research funding: AstraZeneca, Pfizer,

PharmaMar, Roche; travel or accommodation expenses: Amgen, Eli Lilly, Illumina, PharmaMar, Roche.F.d.B.: Consulting: BMS, Pierre Fabre, Mattioli 1885, MCCann Health, MSD, IQVIA, Novartis; Advisory arrangements: Tiziana Life Sciences, BMS, Celgene, Novartis, Servier, Pharm Research Associated, Daiichi Sankyo, Ignyta, Amgen, Pfizer, Octimet Oncology, Incyte, Pierre Fabre, Eli Lilly, Roche, AstraZeneca, Gentili, Dephaforum, Merck & Co., Kenilworth, NJ, Bayer, Fondazione Menarini, Sanofi, Incyte, Taiho; Speaker's fees from BMS, Roche, Merck & Co., Kenilworth, NJ, Bayer, Ignyta, Dephaforum, Biotechspert, Prime Oncology, Pfizer, Nadirex, Ambrosetti, Incyte, Motore Sanità, Events, Fare Comunicazione, Itanet, Nadirex, ESO, AccMed, Idea-z; Grants/contracts: research, unrestricted, and/or travel: Novartis, Roche, BMS, Celgene, Incyte, NMS, the healthcare business of Merck KGaA, Darmstadt, Germany, Kymab, Pfizer, Tesaro, and Merck & Co., Kenilworth, NJ; Traveling expenses from Bristol Myers Squibb, Roche, Celgene, AmgenMN received Travel expenses for meetings from Astrazeneca, speaker honorarium from Accademia della Medicina and Incyte; honoraria from Sandoz, Medpoint SRL and Servier for editorial collaboration. Consultant honoraria from EMD Serono, Basilea Pharmaceutica, Servier, Incyte, MSD Italia, AstraZeneca, Taiho. All other authors declare no conflicts of interest.

### Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41698-025-01231-x>.

**Correspondence** and requests for materials should be addressed to Federico Nichetti.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025