
Article:
Hernandez-Villafuerte, K., Fischer, A. and Latimer, N.R. orcid.org/0000-0001-5304-5585 (2018) Challenges and methodologies in using progression free survival as a surrogate for overall survival In oncology. International Journal of Technology Assessment in Health Care, 34 (3). pp. 300-316. ISSN 0266-4623

https://doi.org/10.1017/S0266462318000338

CHALLENGES AND METHODOLOGIES IN USING PROGRESSION FREE SURVIVAL AS A SURROGATE FOR OVERALL SURVIVAL IN ONCOLOGY

Short title: PFS as a surrogate for OS in Oncology

Karla Hernandez-Villafuerte, PhD. German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, 69120 Heidelberg, Germany. K.hernandezvillafuerte@dkfz.de Previous affiliation: Office of Health Economics. Southside, 7th Floor, 105 Victoria Street, London SW1E 6QT. Tel: +44 (0)20 7747 1412.

Alastair Fischer, PhD. Office of Health Economics. Southside, 7th Floor, 105 Victoria Street, London SW1E 6QT. Tel: +44 (0)20 7747 8886. afischer-fellow@ohe.org

Nicholas Latimer, PhD. ScHARR, University of Sheffield. Regent Court, 30 Regent Street, Sheffield S1 4DA. Tel: +44 (0) 114 222 0821. n.latimer@shef.ac.uk

ABSTRACT

Objectives: A primary outcome in oncology trials is overall survival (OS). However, to estimate OS accurately requires a sufficient number of patients to have died, which may take a long time. If an alternative endpoint is sufficiently highly correlated with OS, it can be used as a surrogate. Progression-free survival (PFS) is the surrogate most often used in oncology, but does not always satisfy the correlation conditions for surrogacy. We analyse the methodologies used when extrapolating from PFS to OS. ***Methods***: Davis, Tappenden (1) previously reviewed the use of surrogate endpoints in oncology, using papers published between 2001 and 2011. We extend this, reviewing papers published between 2012 and 2016. We also examine the reporting of statistical methods to assess the strength of surrogacy. ***Results***: The findings from 2012 to 2016 do not differ substantially from those of 2001 to 2011: the same factors are shown to affect the relationship between PFS and OS. The proportion of papers reporting individual patient data (IPD), strongly recommended for full assessment of surrogacy, remains low: 33%. A wide range of methods has been used to determine the appropriateness of surrogates. While usually adhering to reporting standards, the standard of scholarship appears sometimes to be questionable and the reporting of results often haphazard. ***Conclusion***: Standards of analysis and reporting PFS to OS surrogate studies should be improved by increasing the rigour of statistical reporting and by agreeing to a minimum set of reporting guidelines. Moreover, the use of IPD to assess surrogacy should increase.

Keywords: cancer**,** surrogate endpoints, overall survival (OS), progression-free survival (PFS), post-progression survival (PPS)

FUNDING

The Pharmaceutical Oncology Initiative (POI) of the Association of the British Pharmaceutical Industry (ABPI) commissioned the Office of Health Economics (OHE) in April 2016 to undertake a landscape study on methods and approaches to extrapolation from clinical endpoints measured in trials involving overall survival.

ACKNOWLEDGEMENTS

We gratefully acknowledge the contributions of Francesco Pignatti (EMA), Eli Gavraj and Ian Watson (NICE), Prof Andrew Stevens (Chair of a NICE Appraisal Committee), Oriana Ciani and Prof Rod Taylor (University of Exeter Medical School), and POI member organisations. Nevertheless, views expressed in this paper are those of the authors and are not necessarily those of POI, EMA, NICE or the University of Exeter.

CONFLICTS OF INTEREST

Karla Hernandez-Villafuerte and **Alastair Fischer** do not have any conflicts of interest. **Nicholas Latimer** has acted in a consultancy or advisory role for GlaxoSmithKline, Pfizer, Merck, Sanofi, Astellas, Amgen, AstraZeneca, Janssen, Roche, and Bayer and has received research funding from Novartis, GlaxoSmithKline, Merck and the Pharmacology Oncology Initiative.

INTRODUCTION

Outcomes from clinical and other healthcare trials of most interest to patients and health systems are usually increases in the length and quality of life as a result of treatment. This poses a problem, because to estimate overall survival (OS) sufficiently accurately often requires long-term follow-up. Direct extrapolation of OS to encompass those patients who had not died by the end of the follow-up period may be carried out. Although it is the preferred method of the UK´s National Institute of Health and Care Excellence (NICE), its use is not always satisfactory (2,3). An alternative is to use progression-free survival (PFS) as a surrogate outcome. Trials with an adequate surrogate endpoint can be shorter and involve fewer patients, and can thus help to bring a new drug or treatment to the market sooner, or allow products to be brought to market where the costs of a trial using OS would not be justified by the expected returns. This is beneficial to patients and health systems, and improves returns to manufacturers.

However, these benefits are achieved at the expense of a less accurate measure of final outcome than would occur by waiting for data such as OS to become available. Thus there is a trade-off between time elapsed from the end of study (excluding follow-up) before information becomes available, and the accuracy of the information about the benefits of treatment. This paper analyses the methodologies and challenges faced when using PFS as a surrogate for OS in oncology.

METHODS

Davis, Tappenden (1) conducted a literature review on the use of surrogate endpoints in oncology up to the end of 2011. They identified 266 articles, using citation searching to identify relevant papers from an initial list of three papers already known to the authors. They said a systematic literature review was not feasible, because an exploratory search returned a very large number of references (over 3,000), and because any attempt to make the search more specific resulted in many relevant papers being excluded.

Davis, Tappenden (1) included all reviews that examined a statistical relationship between OS and either PFS or time to progression (TTP) and considered any form of treatment where curing the disease was not expected. Nineteen key articles concerning the relationship between PFS/TTP and OS in advanced/metastatic cancer were included.

We updated the review conducted by Davis, Tappenden (1) to 2016 using similar search methods and selection criteria to preserve as much comparability as possible, and using the 19 papers identified by them as our key papers upon which to base our citation search. We considered only articles in which PFS was mentioned, and excluded those articles that analysed only TTP. Davis, Tappenden (1) did not include radiographic progression-free survival (rPFS) studies, and so neither did we. A previous follow-up of Davis, Tappenden (1) was carried out by Ciani, Davis (4). Our analysis differs from that of Ciani, Davis (4) as our aim has additionally been to examine the statistical methodologies most commonly applied as well as the main challenges faced by authors when assessing the validity of PFS as a surrogate.

In August 2016, we conducted a Google Scholar citation search from January 2012 to June 2016, identifying a total of 790 articles which had cited any of the original 19 key articles identified by Davis, Tappenden (1). We applied four inclusion criteria: 1) mentioned PFS and OS in the title; or 2) mentioned PFS as a surrogate (including surrogate + outcome/endpoint/measure); or 3) analysis of possible surrogate measures in cancer; or 4) analysed endpoints for cancer. An additional seven articles were excluded because they are reviews of previous studies of PFS surrogacy. After applying a series of exclusion criteria (Figure 1), 48 articles were included in the analysis.

The 19 papers reviewed by Davis, Tappenden (1) and the 48 in this paper were mostly summaries of many studies. For each study within a paper, there was a single aggregate data point pair, the average/median PFS and the average/median OS. This pair of points was then used as one observation in an analysis that pooled all such pairs from the many studies considered in the paper. The sample size for the estimated correlations or regression parameters equals the number of studies considered by the paper. Such data

are known as aggregated clinical trial data (ACTD), in which large amounts of data (all the data points of individuals within each trial) have been ignored. In both Davis, Tappenden (1) and our paper, a minority of analyses used unaggregated source data. That is, each patient within the study has a pair of observations which is brought together in a sample, the size of which is the number of patients, for estimating correlations and regression parameters. Such data are known as individual patient data (IPD). Information regarding methodology, data and factors affecting the relationship between PFS and OS were extracted. Author affiliation and publication journal were also collected by KHV. The quality and accuracy of the extraction was verified by AF.

RESULTS

The 2012-2016 results from this review are very similar to those of Davis, Tappenden (1). Davis, Tappenden (1) usually found a positive correlation between PFS/TTP and OS for individual patients, individual trial arms and the treatment effect between trial arms: only 10.5% (2/19) of articles did not support the idea that PFS/TTP could be a useful surrogate for OS. However, the size of the correlation and its statistical significance varied considerably across studies, particularly between cancer types. The authors attributed this variation to the dissimilarities in patient characteristics from study to study, such as tumour type, line of therapy, and diversity of treatment methods.

We classified the results into three groups: 1) papers that explicitly mentioned that PFS is a good surrogate for OS, 2) those that indicated that PFS could be a good surrogate only under certain conditions, and 3) those that concluded that PFS is not a good surrogate for OS. Davis, Tappenden (1) (7/12) and our (17/32) analysis both indicate that around 55% of the articles using ACTD from multiple trials support PFS surrogacy. Eight of the remaining 15 articles in our study do not support surrogacy, while seven articles support PFS surrogacy only under particular conditions (e.g. treatment line).

The lack of IPD is evident if we consider that around 35% of the articles found both by Davis, Tappenden (1) (7/19) and by this review (16/48) include IPD data (Table 1). Among the 10 articles that used solely IPD in our review, four supported surrogacy, 5 did

not and one supported surrogacy for treatments that have a major impact on PFS. Moreover, six of the ten IPD articles in our review were based on information collected in a single Japanese institution, which indicates that the conclusions should be viewed with caution (5-10).

Methodologies and statistical results

In analysing the relationship between PFS and OS, the most usual preference has been for the use of correlation such as Spearman, Pearson or Kendall's τ (71%-34/48) and weighted or unweighted linear regression (73%-35/48) (Table 2), comparable to the findings of Davis, Tappenden (1). Moreover, like Davis, Tappenden (1), we found that many different variations in methodology have been applied which makes it difficult to compare the results of studies (e.g. Aboshi, Kaneko (11) and Bria, Massari (12), Table 2).

Additionally, out of the seven articles that used Pearson correlation, six support surrogacy and one supports surrogacy only for second- or third-line therapy. However, from the 21 articles that used Spearman correlation only seven support surrogacy. This suggests an effect of the correlation test selected. Pearson correlations are affected by outliers more strongly than Spearman correlations; the use of the Pearson correlation without considering outliers could lead to misleading conclusions.

A common practice for testing whether the surrogate is capable of predicting the clinical end point is to analyse the relationship between the actual values of the PFS and OS. Those articles that include ACTD mostly use median PFS and median OS, but in some cases the estimation uses a logarithmic transformation of the variables (27). In the case of IPD articles, the relationship between the actual values per patient of PFS and OS is used to estimate the predictive capacity of PFS.

The effect of treatment in changing PFS to predict the effect of treatment in changing OS is explored throughout the analysis of ACTD. Here it is common to compare hazard ratios

(HRs); however, we also identified articles in which the differences in median PFS and OS were examined (13,33,41).

By comparing the distribution of the observations by Surrogacy (Appropriate surrogate (AP); Depends on particular factors (DPF); PFS is not an appropriate surrogate (NoAP)) (Table 1) with the distribution of the observations split into two depending on the type of data (ACTD, IPD or Both), we observe a significant relationship (Fisher's Exact Test 0.039; Pearson's Chi-squared 0.043) between Surrogacy and Type of data, which suggests that the availability of IPD data affects the final conclusion.

Surrogate threshold effect (STE)

STE is defined as the minimum treatment effect on the surrogate necessary to predict a non-zero effect on the true endpoint (47). It is normally presented as a Hazard Ratio (HR), for instance, a STE equal to 0.8 means that a PFS HR smaller than 0.8 would need to be observed to predict a less than 1.0 HR for OS. This concept has the advantage of not being a yes or no answer to the question of surrogacy, but a lower bound for PFS that if achieved would indicate that a statistically significant effect of the treatment on OS can be predicted. STE has been used relatively more frequently since 2012. In Davis, Tappenden (1), STE was reported in only two of 19 papers; we identified 11 of 48 articles (including five of the six articles that included both IPD and ACTD). However, it is not clear whether STE is affecting authors' final recommendations. For instance, Foster, Renfro (49) and Shi, De Gramont (50) supported PFS surrogacy with an STE for PFS HR smaller than 0.70 while Ciani, Buyse (41) rejected surrogacy with an STE of 0.80 (Table 3).

Weaknesses of the current approaches

1) Lack of rigour in applying methodology.

Weighted linear regression, the most frequently-used method of analysis, is based on assumptions that are not tested in the majority of articles when analysing surrogacy. In only a few such cases was the type of model mentioned (14). Exceptions are Félix, Aragão (16) that use the Generalized Method of Moments (GMM) to control for

heteroscedasticity; and Johnson, Liauw (35) whose results showed unsatisfactory diagnostics with non-normality and heteroscedasticity in the residuals.

Although linear regression assumptions should be so widely known as to be considered irrelevant to report them, their absence leaves many analyses open to the suggestion of failure to handle complications, such as the presence of outliers. Clear outliers are shown in Yoshino, Imai (9) and Moriwaki, Yamamoto (38). Only 23% (11/48) of studies consider or mention outliers. In five of the 11 cases authors test the sensitivity of the results by applying a 'leave-one-out' strategy (26,47,49,52). The six remaining articles test the sensitivity of the results by excluding those trials that are considered outliers (13,16,23,24,32,36).

Publication bias could also have a significant impact on the results, particularly for ACTD. Out of the 31 articles that include systematic literature reviews, 11 mentioned publication bias as a possible limitation of the study while an additional six articles included a step to overcome the possible bias. Among the six, some articles considered both published and unpublished clinical trials (15,30,40); the others analysed the extent to which bias represents a problem using Egger's regression test (13,24,41).

2) Apparently inconsistent conclusions

The German Institute of Quality and Efficiency in Health Care (IQWiG) framework considers 'low correlation' to be present when the upper limit (95% confidence interval) of the correlation is under 0.70 (Pearson R^2 smaller than 0.49) (53). Two out of the five articles concerning colorectal cancer that support surrogacy have correlation values lower than 0.70 (19,50). For renal carcinoma, Delea, Khuu (13) and Halabi, Rini (44) support PFS as a surrogate despite values of association lower than 0.7 (Table 2).

Petrelli and Barni (20) stand out because of the lack of consistency between supporting PFS surrogacy and having results in which R^2 (equal to 0.00) and correlation values (treatment effect correlation 0.59 and actual values correlation 0.26) are particularly low. They observed a weak correlation between PFS and OS for NSCLC. However, they still supported PFS as a surrogate for OS, a decision that appears to be influenced by the

slope of the linear regression. The slope suggests that a one-month gain in PFS will be linked to three weeks' prolongation in OS. However, the reasons for concluding that the surrogacy is supported remain unclear.

Table 2 shows inconsistencies between IPD and aggregate approaches (40,42). Foster, Renfro (49), based on ACTD, supported surrogacy, while IPD data from the same trial were less conclusive.

Finally, it is possible that different studies analyse PFS surrogacy for a particular cancer type by including a ^{similar} list of clinical trials. However, it is not possible to observe whether similar lists of clinical trials could result in different conclusions since the references of the included trials are not mentioned in some of the papers (14).

Challenges for analysing PFS as a surrogate of OS

Based on the variables included as part of sensitivity analyses or that have been included in multivariate analyses in the papers included in this review, we identify a group of factors that could affect the relationship between PFS and OS and appear in at least five studies.

1) Type of treatment and/or therapy (11-22,26,32,33,35,36,38,40-42,47,49,50)

The literature suggests that the relationship between PFS and OS can be different within the same cancer trial depending on the treatment applied or the therapy selected.

2) Treatment line (14,21,22,25,26,28,33,34,36,40,41,49,52)

In some cases the analysis cannot validate the surrogacy for first line therapy, as distinct from the second or third line therapy, mainly because post-progression survival obscures the results of the first line treatments. Petrelli and Barni (32), who analysed 20 first line clinical trials, proposed that the decreases in correlation between PFS HR and OS HR observed in recent years is likely to be due to the influence of post-progression treatments.

3) Year of the trial (11,13,15,16,18,19,22,28,31-34,40)

The importance of the year in which the clinical trial was conducted or published was explained by the number of drugs available having increased (11 and because the

criteria applied to measure progression have changed (e.g. RECIST published in 2000 was modified in 2010 to mRECIST (54)).

4) Sub-group of patients or tumour type (10,11,16,18,21,26,32,34,39,52)

As in Davis, Tappenden (1), the results from the validation of PFS as a surrogate point for OS vary substantially between cancer types. Six out of 16 articles for lung cancer conclude that PFS is not an appropriate surrogate for OS, and consistency does not improve when we consider the line of therapy and the phase of the clinical trial. This might be related to the fact that the criteria for supporting surrogacy differ considerably between studies. Additionally, it might not be just the observed relationship that is changing between studies. This suggests a need to standardise criteria.

5) Definition of PFS and other measures

Disease progression is often defined differently between clinical trials. This heterogeneity pertains to the period within which patients are evaluated; time intervals between radiological and clinical assessments; and what constitutes patient progression (e.g. variation of the size of the tumour, and tumour characteristics). There are many other forms of heterogeneity, some of which have been mentioned in sub-sections 1) to 4) above. The inclusion of Phase II trials is also likely to increase heterogenity. This occurs because at most one arm of a dose-determining trial should be featured in a subsequent Phase III trial. Arms in which patients are over-dosed and on average die sooner than those in the control group should not feature in Phase III trials. Their inclusion in papers that measure the correlation between PFS and OS will reduce the estimated correlation.

Twenty two articles mentioned the problem of heterogeneity as a limitation, but did not adjust the methodology in response to the problem. A further seven studies adjusted the methodology (18,23,26,31-33,42), two of which included only clinical trials that have the same set of progression criteria (RECIST criteria) (23,42). Three of the seven included variables such as presence of measurable lesions and tumour response in sensitivity analysis (18,31,33). Finally, two of the seven studies used established definitions to

extract the information collected from the clinical trials, regardless of the terminology used by the original authors (26,32).

Additionally, 19 out of the 32 studies based on trial data combined PFS and TTP into a single surrogate measure. In addition to progression, PFS includes death as a result of any cause while in the case of TTP the event of interest is only disease progression, although some authors consider, as part of TTP, deaths caused by the disease in question (e.g. Burzykowski, Buyse (55) identified by Davis, Tappenden (1)). All-cause mortality can dilute the association between PFS/TTP and OS. Nine of the 19 articles that include PFS and TTP analyse the sensitivity of the results by breaking down the articles into those that measure PFS and those that measure TTP (13,15,22,25,31-33,38,42). In contrast to what we would expect, Delea, Khuu (13), Petrelli and Barni (32) and Shitara, Ikeda (15) found that studies that include PFS have a higher correlation with OS compared with studies that include TTP. However, Moriwaki, Yamamoto (38) found a slightly lower correlation when TTP trials were excluded.

6) Geographical context (11,14,15,28,33,40)

A reason given to explain geographical differences in trial results is the variation in comparator (i.e. standard) treatments between Asian and occidental countries. In addition, in advanced gastric cancer, Shitara, Ikeda (15) pointed out a number of differences in tumour characteristics and practice patterns (e.g. surgery and chemotherapy) that have been identified between Asian and occidental countries.

7) Crossover (13,23,30,33,41)

Some clinical trials allow crossover to the experimental regimen upon disease progression. This hinders the analysis of the treatment effect on OS. Eighteen out of the 32 articles that use ACTD mentioned crossover while six articles considered it during the estimation. For renal cell carcinoma, Delea, Khuu (13) indicate that the link between the effect of the treatment on PFS and the effect of the treatment on OS was stronger in studies that did not allow crossover. In melanoma, Flaherty, Hennig (23) suggested that correlation coefficients for the nine trials without crossover were significant and more

than 7 percentage points higher than with crossover. Hotta, Suzuki (30), in studying NSCLC, suggest that for clinical trials in which the median proportion of crossover was lower than 1%, the association between the HRs of PFS and OS was strong. Kim and Prasad (56), identified by Davis, Tappenden (1), evaluated previous publications to assess the strength of the surrogate-survival correlation among cancer drugs approved. They found no significant differences in survival benefit between clinical trials with or without crossover. They suggest that the results are opposed to the commonly-shared idea that crossover masks OS benefits, possibly because crossover prevents observation of late toxicity. Contrary to other studies, in an analysis of colorectal cancer trials, Adunlin, Cyrus (33) found that among crossover trials the strength of the association between PFS and OS was higher.

Characteristics of post-progression survival

Characteristics of post-progression survival (treatment line $1^{st}/2^{nd}/3^{rd}$, year of the clinical trial, crossover between control and treatment arms, newly diagnosed vs recurrent, and sub-group of patients) and the fact that an important number of articles analysed post-progression survival (PPS) together with PFS suggest that PPS has a role in the discussion of the validation of PFS as a surrogate of OS. Amir, Seruga (29) indicate that when PPS is short, the correlation between OS and PFS is higher than when PPS is long. For patients with advanced NSCLC, Suzuki, Hirashima (37) identified the optimal point of correlation of the HR for PFS and the HR for OS by analysing every 1 month of PPS. They found that the correlation between the HR for PFS and for OS increases for a PPS of less than 6 months and then decreases (<4 months 0.70; <6 months 0.77; <9 months 0.46). From the 16 articles that analysed PPS, 13 suggest that the relationship between OS and PPS is stronger than between OS and PFS and one of the remaining three pointed out a high correlation between PFS and PPS.

A group of Japanese researchers specialising in a study of factors that affect the relationship between PPS and OS (5-10) suggests that the significant factors to explain the effect of PPS on OS are:

- Number of regimens employed after progression
- Response to the second or third-line treatment
- Performance status at progression
- PFS of first line chemotherapy
- Tumour stage after initial treatment
- The presence of distant metastases at recurrence

DISCUSSION

The percentage of articles that conclude that PFS is an appropriate surrogate for OS (52%) is higher than the percentage of those that do not support surrogacy (25%). An additional 23% of the samples suggest that surrogacy depends on factors such as the length of the PPS and whether the treatment was first or subsequent line. In such a complicated area, it is no wonder that simple rules of thumb, to determine whether a surrogate endpoint can replace OS, will not work in all situations. Additionally, it seems that different investigators use different rules of thumb in the same circumstances.

The first set of criteria to establish whether a surrogate would be an adequate replacement for OS was proposed by Prentice (57). His criteria have been amended and elaborated since then. Ciani, Davis (4) summarised three different frameworks that are currently applied to validate the strength of the evidence (IQWiG, Biomarker-Surrogacy Evaluation Schema, and Elston and Taylor's frameworks). All include the Prentice (57) criteria, but also analyse factors that could influence the strength of the relationship, such as the quality of the data and characteristics of the clinical trial.

Ciani, Buyse (58) highlight three further conditions. First, the strength of the association between the surrogate endpoint and the final outcome should be measured through approaches such as regression and meta-analysis. Second, it is necessary that the effect on the final outcome can be predicted and quantified based on the effect on the surrogate. The effect of the treatment on PFS must be large enough to predict an

improvement in OS. Third, the level of evidence supporting the relationship between the surrogate endpoint and the desired outcome needs to be considered. A strong correlation should be observed between the surrogate and the end point based on individual patient data as well as between the treatment effect on the surrogate end point and the final outcomes across multiple randomised trials. Similarly, Buyse, Molenberghs (59) propose that to validate a surrogate endpoint, it is necessary to analyse both individual and trial level data. ACTD is important for testing the relationship between the treatment effect on PFS and the treatment effect on OS while the IPD allows the analysis of the relationship between the actual value of PFS and the actual value of OS. Thus, despite the existence of recognised surrogate validation criteria, which have been developed over time, no consistent application of these criteria was observed in the studies we reviewed – different authors made different, often arbitrary assessments of surrogate adequacy.

The lack of any substantial proportionate increase in the number of articles including IPD data between Davis, Tappenden (1) study and our analysis suggests the rate of progress in this field is being hampered by an unwillingness by most pharmaceutical companies to provide IPD data to independent and well-qualified researchers for analysis or even to report analyses based on IPD when those data have routinely been collected. Nevertheless, some progress in the topic has been observed, e.g. firms have joined recent initiatives such as clinicalstudydatarequest.com or project data sphere that allows researchers to analyse pooled IPD data sets.

The existence of heterogeneity in the definition of progression among clinical trials and a lack of clear information in the clinical trial reports as to how disease progression was evaluated (15,17,40) indicate that there is a need to standardise clinical trial protocols to provide comparability between trials for the same cancer type.

Finally, our search process found additional evidence that went beyond the scope of Davis, Tappenden (1) analysis. Stevens, Philipson (60) outlined an economic approach which bears on both clinical effectiveness and on cost effectiveness, suggesting a

framework for factoring the use of surrogates into the decision-making process. Perhaps surprisingly, the benefits (or costs) of earlier adoption of a new technology that the use of a surrogate endpoint will usually allow are not taken into consideration when assessing new treatments. The longer the lag between the results of a trial using a surrogate endpoint rather than OS, the greater the additional benefits of using a surrogate should be, provided the surrogate is valid and that subsequent treatments do not act as confounders. However, as Davis Tappenden (1) explain, even when strong consistent evidence supporting a correlation between the treatment effects is available, it is unclear how that should be converted into a quantified relationship between PFS and OS treatment effects within a cost-effectiveness model.

CONCLUSION

The analysis strongly suggests that the use of IPD to assess surrogacy should increase. A case could be made for release of all IPD as a condition of publication. As in Davis, Tappenden (1), our findings show that the availability of such information has been limited, though recent data-sharing initiatives may be changing that.

There is a high variation in the characteristics of the methodologies and little apparent consistency in what should be considered appropriate statistical estimation methodology. Thus the need for standardisation that allows for more consistent results. Standardization, in the form of adhering to common definitions, statistical techniques and a checklist of necessary items in reporting results, would often be virtually costless. This could facilitate the use of PFS by policy-makers -- if it were deemed appropriate -- based upon standardised validation methodology, and could increase both the speed and accuracy of their decision-making.

Many of the factors that affect the validation of surrogacy are related to the length and characteristics of post-progression survival. Procedures for gathering information on factors affecting the post-progression management of a disease should be described in protocols for following-up clinical trial patients, making it possible to derive stronger conclusions from statistical analysis.

Some limitations of the study need to be mentioned. First, it is not a full literature review. We conducted a citation search based on Davis, Tappenden (1) 19 studies that we assume captured all relevant articles. Discussions with experts and comparisons with previous systematic reviews suggest that no relevant article has been excluded from the analysis. Second, this is also not a systematic literature review of any particular cancer type. Therefore, analysing whether PFS should or should not be used in any particular case was outside of the scope of this analysis. It is recommended that the factors that affect the relationship between PFS and OS by cancer type should be analysed in order to understand the particular challenges faced in each case. Third, for pragmatic reasons, our exclusion of TTP ignores the possibility that the names of TTP and PFS have in error been used interchangeably (61).

Finally, in addition to using one of the frameworks promoted by Ciani, Davis (4) to ensure a higher standard of validation of the strength of evidence, both researchers and policy-makers in an area that makes use of surrogate endpoints need to be aware that the statistical methodology must be properly understood and documented. The importance that validating PFS as a surrogate for OS may have on allowing patients to access new health technologies more quickly should not be undermined by a poor knowledge of the methodology applied. The results of this study are broadly in line with those of Kemp and Prasad (62) who have concluded that the use of surrogate outcomes should be limited to situations where a surrogate has demonstrated robust ability to predict meaningful benefits, or where cases are dire, rare or with few treatment options.

REFERENCES

1. Davis S, Tappenden P, Cantrell A. A review of studies examining the relationship between progression-free survival and overall survival in advanced or metastatic cancer. London: National Institute for Health and Care Excellence (NICE), 2012.

2. Latimer NR. Survival analysis for economic evaluations alongside clinical trials: Extrapolation with patient-level data. Medical Decision Making. 2013;33(6):743-54.

3. Latimer NR, Abrams KR, Lambert PC, Crowther MJ, Wailoo AJ, Morden JP, et al. Adjusting survival time estimates to account for treatment switching in randomized controlled trials: An economic evaluation context. Medical Decision Making. 2014;34(3):387-402.

4. Ciani O, Davis S, Tappenden P, Garside R, Stein K, Cantrell A, et al. Validation of surrogate endpoints in advanced solid tumors: Systematic review of statistical methods, results, and implications for policy makers. International Journal of Technology Assessment in Health Care. 2014;30(3):312-24.

53. IQWiG. Validity of surrogate endpoints in oncology: Executive summary of rapid report A10-05, Version 1.1. Institute for Quality and Efficiency in Health Care: Executive Summaries. Cologne, Germany: Institute for Quality and Efficiency in Health Care (IQWiG); 2005.

54. Lencioni R, Llovet JM. Modified RECIST (mRECIST) assessment for hepatocellular carcinoma. Seminars in Liver Disease. 2010;30(1):52-60.

55. Burzykowski T, Buyse M, Piccart-Gebhart MJ, Sledge G, Carmichael J, Lück H-J, et al. Evaluation of tumor response, disease control, progression-free survival, and time to progression as potential surrogate end points in metastatic breast cancer. Journal of Clinical Oncology. 2008;26(12):1987-92.

56. Kim C, Prasad V. Strength of validation for surrogate end points used in the US Food and Drug Administration's approval of oncology drugs. Mayo Clinic Proceedings. 2016;91(6):713-25.

57. Prentice RL. Surrogate endpoints in clinical trials: Definition and operational criteria. Statistics in Medicine. 1989;8(4):431-40.

58. Ciani O, Buyse M, Drummond M, Rasi G, Saad ED, Taylor RS. Use of surrogate end points in healthcare policy: A proposal for adoption of a validation framework. Nature Reviews Drug Discovery. 2016;15(7):516.

59. Buyse M, Molenberghs G, Burzykowski T, Renard D, Geys H. The validation of surrogate endpoints in meta-analyses of randomized experiments. Biostatistics. 2000;1(1):49-67.

60. Stevens W, Philipson T, Wu Y, Chen C, Lakdawalla D. A cost-benefit analysis of using evidence of effectiveness in terms of progression free survival in making reimbursement decisions on new cancer therapies. Forum for Health Economics and Policy. 2014;17(1):21-52.

61. Saad ED, Katz A. Progression-free survival and time to progression as primary end points in advanced breast cancer: Often used, sometimes loosely defined. Annals of Oncology. 2008; 20(3): 460-64.

62. Kemp R, Prasad V. Surrogate endpoints in oncology: when are they acceptable for regulatory and clinical decisions, and are they currently overused?. BMC Medicine. 2017; 15(1):134.

5. Imai H, Mori K, Ono A, Akamatsu H, Taira T, Kenmotsu H, et al. Individual-level data on the relationships of progression-free survival and post-progression survival with overall survival in patients with advanced non-squamous non-small cell lung cancer patients who received second-line chemotherapy. Medical Oncology. 2014;31(8):1-7.

6. Imai H, Mori K, Wakuda K, Ono A, Akamatsu H, Shukuya T, et al. Progression-free survival, post-progression survival, and tumor response as surrogate markers for overall survival in patients with extensive small cell lung cancer. Annals of Thoracic Medicine. 2015;10(1):61-6.

7. Kasahara N, Imai H, Kaira K, Mori K, Wakuda K, Ono A, et al. Clinical impact of post-progression survival on overall survival in patients with limited-stage disease small cell lung cancer after first-line chemoradiotherapy. Radiology and Oncology. 2015;49(4):409-15.

8. Yoshino R, Imai H, Mori K, Takei K, Tomizawa M, Kaira K, et al. Surrogate endpoints for overall survival in advanced non-small-cell lung cancer patients with mutations of the epidermal growth factor receptor gene. Molecular and Clinical Oncology. 2014;2(5):731-6.

9. Yoshino R, Imai H, Mori K, Tomizawa Y, Takei K, Tomizawa M, et al. Clinical impact of postprogression survival for overall survival in elderly patients (aged 75 years or older) with advanced nonsmall cell lung cancer. Journal of Cancer Research and Therapeutics. 2015;11(3):606.

10. Shitara K, Matsuo K, Muro K, Doi T, Ohtsu A. Progression-free survival and post-progression survival in patients with advanced gastric cancer treated with first-line chemotherapy. Journal of Cancer Research and Clinical Oncology. 2013;139(8):1383-9.

11. Aboshi M, Kaneko M, Narukawa M. Factors affecting the association between overall survival and progression-free survival in clinical trials of first-line treatment for patients with advanced non-small cell lung cancer. Journal of Cancer Research and Clinical Oncology. 2014;140(5):839-48.

12. Bria E, Massari F, Maines F, Pilotto S, Bonomi M, Porta C, et al. Progression-free survival as primary endpoint in randomized clinical trials of targeted agents for advanced renal cell carcinoma: Correlation with overall survival, benchmarking and power analysis. Critical Reviews in Oncology/Hematology. 2015;93(1):50-9.

13. Delea TE, Khuu A, Heng DY, Haas T, Soulieres D. Association between treatment effects on disease progression end points and overall survival in clinical studies of patients with metastatic renal cell carcinoma. British Journal of Cancer. 2012;107(7):1059-68.

14. Li X, Liu S, Gu H, Wang D. Surrogate end points for survival in the target treatment of advanced non-small-cell lung cancer with gefitinib or erlotinib. Journal of Cancer Research and Clinical Oncology. 2012;138(11):1963-9.

15. Shitara K, Ikeda J, Yokota T, Takahari D, Ura T, Muro K, et al. Progression-free survival and time to progression as surrogate markers of overall survival in patients with advanced gastric cancer: Analysis of 36 randomized trials. Investigational New Drugs. 2012;30(3):1224-31.

16. Félix J, Aragão F, Almeida JM, Calado FJ, Ferreira D, Parreira AB, et al. Time-dependent endpoints as predictors of overall survival in multiple myeloma. BMC Cancer. 2013;13(122):1-12.

17. Giessen C, Laubender RP, Ankerst DP, Stintzing S, Modest DP, Mansmann U, et al. Progression-free survival as a surrogate endpoint for median overall survival in metastatic colorectal cancer: Literature-based analysis from 50 randomized first-line trials. Clinical Cancer Research. 2013;19(1):225-35.

18. Han K, Ren M, Wick W, Abrey L, Das A, Jin J, et al. Progression-free survival as a surrogate endpoint for overall survival in glioblastoma: A literature-based meta-analysis from 91 trials. Neuro-Oncology. 2013;16(5):696-706.

19. Petrelli F, Barni S. Correlation of progression-free and post-progression survival with overall survival in advanced colorectal cancer. Annals of Oncology. 2013;24(1):186-92.

20. Petrelli F, Barni S. Is overall survival still the primary endpoint in maintenance non-small cell lung cancer studies? An analysis of phase III randomised trials. Translational Lung Cancer Research. 2013;2(1):6-13.

21. Sidhu R, Rong A, Dahlberg S. Evaluation of progression-free survival as a surrogate endpoint for survival in chemotherapy and targeted agent metastatic colorectal cancer trials. Clinical Cancer Research. 2013;19(5):969-76.

22. Beauchemin C, Cooper D, Lapierre M-È, Yelle L, Lachaine J. Progression-free survival as a potential surrogate for overall survival in metastatic breast cancer. OncoTargets and Therapy. 2014;7(1):1101-10.

23. Flaherty KT, Hennig M, Lee SJ, Ascierto PA, Dummer R, Eggermont AMM, et al. Surrogate endpoints for overall survival in metastatic melanoma: A meta-analysis of randomised controlled trials. The Lancet Oncology. 2014;15(3):297-304.

24. Singh S, Wang X, Law C. Association between time to disease progression end points and overall survival in patients with neuroendocrine tumors. Gastrointestinal Cancer: Targets and Therapy 2014;4(1):103-13.

25. Cartier S, Zhang B, Rosen VM, Zarotsky V, Bartlett JB, Mukhopadhyay P, et al. Relationship between Treatment Effects on Progression-Free Survival and Overall Survival in Multiple Myeloma: A Systematic Review and Meta-Analysis of Published Clinical Trial Data. Oncology Research and Treatment. 2015;38(3):88-94.

26. Chen Y-P, Sun Y, Chen L, Mao Y-P, Tang L-L, Li W-F, et al. Surrogate endpoints for overall survival in combined chemotherapy and radiotherapy trials in nasopharyngeal carcinoma: Meta-analysis of randomised controlled trials. Radiotherapy and Oncology. 2015;116(2):157-66.

27. Giessen C, Laubender RP, Ankerst DP, Stintzing S, Modest DP, Schulz C, et al. Surrogate endpoints in second-line treatment for mCRC: A systematic literature-based analysis from 23 randomised trials. Acta Oncologica. 2015;54(2):187-93.

28. Petrelli F, Coinu A, Borgonovo K, Cabiddu M, Barni S. Progression-free survival as surrogate endpoint in advanced pancreatic cancer: Meta-analysis of 30 randomized first-line trials. Hepatobiliary & Pancreatic Diseases International. 2015;14(2):124-31.

29. Amir E, Seruga B, Kwong R, Tannock IF, Ocaña A. Poor correlation between progression-free and overall survival in modern clinical trials: Are composite endpoints the answer? European Journal of Cancer. 2012;48(3):385-8.

30. Hotta K, Suzuki E, Di Maio M, Chiodini P, Fujiwara Y, Takigawa N, et al. Progression-free survival and overall survival in phase III trials of molecular-targeted agents in advanced non-small-cell lung cancer. Lung Cancer. 2013;79(1):20-6.

31. Kawakami H, Okamoto I, Hayashi H, Taguri M, Morita S, Nakagawa K. Postprogression survival for first-line chemotherapy in patients with advanced gastric cancer. European Journal of Cancer. 2013;49(14):3003-9.

32. Petrelli F, Barni S. Surrogate endpoints in metastatic breast cancer treated with targeted therapies: an analysis of the first-line phase III trials. Medical Oncology. 2014;31(1):1-8.

33. Adunlin G, Cyrus JWW, Dranitsaris G. Correlation between progression-free survival and overall survival in metastatic breast cancer patients receiving anthracyclines, taxanes, or targeted therapies: A trial-level meta-analysis. Breast Cancer Research and Treatment. 2015;154(3):591-608.

34. Hotta K, Kato Y, Leighl N, Takigawa N, Gaafar RM, Kayatani H, et al. Magnitude of the benefit of progression-free survival as a potential surrogate marker in phase 3 trials assessing targeted agents in molecularly selected patients with advanced non-small cell lung cancer: Systematic review. PloS One. 2015;10(3):e0121211.

35. Johnson KR, Liauw W, Lassere MND. Evaluating surrogacy metrics and investigating approval decisions of progression-free survival (PFS) in metastatic renal cell cancer: A systematic review. Annals of Oncology. 2015;26(3):485-96.

36. Özer-Stillman I, Strand L, Chang J, Mohamed AF, Tranbarger-Freier KE. Meta-analysis for the association between overall survival and progression-free survival in gastrointestinal stromal tumor. Clinical Cancer Research. 2015;21(2):295-302.

37. Suzuki H, Hirashima T, Okamoto N, Yamadori T, Tamiya M, Morishita N, et al. Relationship between progression-free survival and overall survival in patients with advanced non-small cell lung cancer treated with anticancer agents after first-line treatment failure. Asia-Pacific Journal of Clinical Oncology. 2015;11(2):121-8.

38. Moriwaki T, Yamamoto Y, Gosho M, Kobayashi M, Sugaya A, Yamada T, et al. Correlations of survival with progression-free survival, response rate, and disease control rate in advanced biliary tract cancer: A meta-analysis of randomised trials of first-line chemotherapy. British Journal of Cancer. 2016;114(8):881-8.

39. Petrelli F, Barni S. Surrogate end points and postprogression survival in renal cell carcinoma: An analysis of first-line trials with targeted therapies. Clinical Genitourinary Cancer. 2013;11(4):385-9.

40. Shitara K, Matsuo K, Muro K, Doi T, Ohtsu A. Correlation between overall survival and other endpoints in clinical trials of second-line chemotherapy for patients with advanced gastric cancer. Gastric Cancer. 2014;17(2):362-70.

41. Ciani O, Buyse M, Garside R, Peters J, Saad ED, Stein K, et al. Meta-analyses of randomized controlled trials show suboptimal validity of surrogate outcomes for overall survival in advanced colorectal cancer. Journal of Clinical Epidemiology. 2015;68(7):833-42.

42. Terashima T, Yamashita T, Takata N, Nakagawa H, Toyama T, Arai K, et al. Post-progression survival and progression-free survival in patients with advanced hepatocellular carcinoma treated by sorafenib. Hepatology Research. 2015;46(7):650-6.

43. Galsky MD, Krege S, Lin CC, Hahn N, Ecke T, Moshier E, et al. Relationship between 6-and 9-month progression-free survival and overall survival in patients with metastatic urothelial cancer treated with first-line cisplatin-based chemotherapy. Cancer. 2013;119(16):3020-6.

44. Halabi S, Rini B, Escudier B, Stadler WM, Small EJ. Progression-free survival as a surrogate endpoint of overall survival in patients with metastatic renal cell carcinoma. Cancer. 2014;120(1):52-60.

45. Négrier S, Bushmakin AG, Cappelleri JC, Korytowsky B, Sandin R, Charbonneau C, et al. Assessment of progression-free survival as a surrogate end-point for overall survival in patients with metastatic renal cell carcinoma. European Journal of Cancer. 2014;50(10):1766-71.

46. Laporte S, Squifflet P, Baroux N, Fossella F, Georgoulias V, Pujol J-L, et al. Prediction of survival benefits from progression-free survival benefits in advanced non-small-cell lung cancer: Evidence from a meta-analysis of 2334 patients from 5 randomised trials. BMJ Open. 2013;3(3):e001802.

47. Mauguen A, Pignon J-P, Burdett S, Domerg C, Fisher D, Paulus R, et al. Surrogate endpoints for overall survival in chemotherapy and radiotherapy trials in operable and locally advanced lung cancer: A re-analysis of meta-analyses of individual patients' data. The Lancet Oncology. 2013;14(7):619-26.

48. Agarwal N, Bellmunt J, Maughan BL, Boucher KM, Choueiri TK, Qu AQ, et al. Six-month progression-free survival as the primary endpoint to evaluate the activity of new agents as second-line therapy for advanced urothelial carcinoma. Clinical Genitourinary Cancer. 2014;12(2):130-7.

49. Foster NR, Renfro LA, Schild SE, Redman MW, Wang XF, Dahlberg SE, et al. Multitrial evaluation of progression-free survival as a surrogate end point for overall survival in first-line extensive-stage small-cell lung cancer. Journal of Thoracic Oncology. 2015;10(7):1099-106.

50. Shi Q, De Gramont A, Grothey A, Zalcberg J, Chibaudel B, Schmoll H-J, et al. Individual patient data analysis of progression-free survival versus overall survival as a first-line end point for metastatic colorectal cancer in modern randomized trials: Findings from the analysis and research in cancers of the digestive system database. Journal of Clinical Oncology. 2015;33(1):22-8.

51. Paoletti X, Oba K, Bang Y-J, Bleiberg H, Boku N, Bouché O, et al. Progression-free survival as a surrogate for overall survival in advanced/recurrent gastric cancer trials: A meta-analysis. Journal of the National Cancer Institute. 2013;105(21):1667-70.

52. Michiels S, Pugliano L, Marguet S, Grun D, Barinoff J, Cameron D, et al. Progression-free survival as surrogate endpoint for overall survival in clinical trials of HER2-targeted

agents in HER2-positive metastatic breast cancer. Annals of Oncology. 2016;27(6):1029-34.