



On the relationship between tumour growth rate and survival in non-small cell lung cancer

Hitesh B. Mistry

Division of Pharmacy, University of Manchester, Manchester, United Kingdom

ABSTRACT

A recurrent question within oncology drug development is predicting phase III outcome for a new treatment using early clinical data. One approach to tackle this problem has been to derive metrics from mathematical models that describe tumour size dynamics termed re-growth rate and time to tumour re-growth. They have shown to be strong predictors of overall survival in numerous studies but there is debate about how these metrics are derived and if they are more predictive than empirical end-points. This work explores the issues raised in using model-derived metric as predictors for survival analyses. Re-growth rate and time to tumour re-growth were calculated for three large clinical studies by forward and reverse alignment. The latter involves re-aligning patients to their time of progression. Hence, it accounts for the time taken to estimate re-growth rate and time to tumour re-growth but also assesses if these predictors correlate to survival from the time of progression. I found that neither re-growth rate nor time to tumour re-growth correlated to survival using reverse alignment. This suggests that the dynamics of tumours up until disease progression has no relationship to survival post progression. For prediction of a phase III trial I found the metrics performed no better than empirical end-points. These results highlight that care must be taken when relating dynamics of tumour imaging to survival and that bench-marking new approaches to existing ones is essential.

Submitted 6 September 2017
Accepted 9 November 2017
Published 29 November 2017

Corresponding author
Hitesh B. Mistry,
hitesh.mistry@manchester.ac.uk

Academic editor
Chris Brown

Additional Information and
Declarations can be found on
page 10

DOI 10.7717/peerj.4111

© Copyright
2017 Mistry

Distributed under
Creative Commons CC-BY 4.0

OPEN ACCESS

Subjects Mathematical Biology, Clinical Trials, Oncology, Pharmacology, Statistics

Keywords Survival, Longitudinal imaging, Lung cancer

INTRODUCTION

A key question being posed in early oncology drug development at the end of phase II (and increasingly at the end of phase I expansion studies) is: given past imaging information, what is the overall survival (OS) outcome in a phase III study likely to be? This question has been primarily addressed by analysing efficacy data using RECIST criteria (Response Evaluation Criteria In Solid Tumours) (*Therasse et al., 2000; Eisenhauer et al., 2009*). RECIST v1.0 and v1.1 involves the evaluation of a drug's efficacy against both target and non-target lesions. The drug's effect on target lesions are reported via the Sum of Longest Diameters (SLD) marker which is a continuous variable whereas drug effect on non-target lesions is reported via qualitative descriptions: increase, decrease or no-change. Both are recorded over time as is the occurrence of a new lesion. The variables discussed above can also be combined to produce one variable, RECIST response classification which consists

of four categories: Complete Response (CR), Partial Response (PR), Stable Disease (SD) or Progressive Disease (PD). CR represents complete disappearance of both target and non-target lesions. PR and SD are simply categorised versions of percentage change in SLD. PD as well as being a categorised version of percentage change in SLD also considers increase in non-target lesion size and occurrence of a new lesion. An important continuous metric of interest is progression free survival (PFS) time: the time from treatment initiation to PD or death. This time-point is of importance for two reasons; firstly it marks the end of imaging data collection and secondly, it also signals the end of the current treatment and the beginning of the next. Therefore, the imaging data collected can only provide one with information on the effect of the first treatment given during the trial. Without knowing what effect the treatment post progression has on the disease clearly makes survival prediction challenging. However, if one assumes the subsequent treatments have minimal effect on disease, compared to initial treatment, then survival prediction may be possible. The debate then moves onto how best to analyse the imaging data collected up until disease progression.

The option of using continuous measures involving changes in SLD over the categorisation scheme has been a long-standing debate when analysing early patient response to treatment ([An et al., 2011](#); [An et al., 2015](#); [Toffart et al., 2014](#)). These works have shown that there was no benefit in terms of survival concordance probability when using continuous changes in SLD compared to the current RECIST response classification. More elaborated approaches have therefore explored the use of mathematical models to fit the time-series of SLD ([Stein et al., 2008](#); [Stein et al., 2009](#); [Stein et al., 2012](#); [Wang et al., 2009](#); [Blagoev et al., 2013](#); [Claret et al., 2013](#); [Bruno, Mercier & Claret, 2014](#); [Han et al., 2016](#)). [Figure 1A](#) shows the metrics that have been derived from these models of SLD time-series. These are the decay rate (DR), the re-growth rate (GR), the model estimated percentage change at a certain time-point and the time to tumour re-growth (TTG). In particular, TTG and GR have been claimed to be significant predictors of survival ([Stein et al., 2008](#); [Stein et al., 2009](#); [Stein et al., 2012](#); [Blagoev et al., 2013](#); [Claret et al., 2013](#); [Han et al., 2016](#)) when calculated using an exponential decay/growth model. However, concerns about how these two metrics are estimated have been raised. The first relates to drop-outs not being accounted for through a joint longitudinal and survival model ([Ibrahim, Chu & Chen, 2010](#); [Mansmann & Laubender, 2013](#)). The second concerns the time-dependent bias of TTG and GR, that is the time taken to estimate these quantities was never accounted for in those previous studies ([Van Walraven et al., 2004](#); [Mistry, 2016](#)).

Contradictory results have been reported regarding the predictive power of these metrics. In one hand, [Sharma et al. \(2015\)](#) used TTG within a resampling of a single phase III study to assess phase II trial design and endpoints. They showed that TTG was equal if not superior to PFS in making the correct decision with regards to moving from phase II to III given an 18 week landmark time-point. On the other hand, [Kaiser \(2013\)](#) used a larger study and found that PFS was a superior endpoint compared to GR in five of his six trials. There are two possible reasons why the results from Sharma et al. and Kaiser differ: (i) the difference in PFS between treatments, in the studies used by Kaiser, were greater than

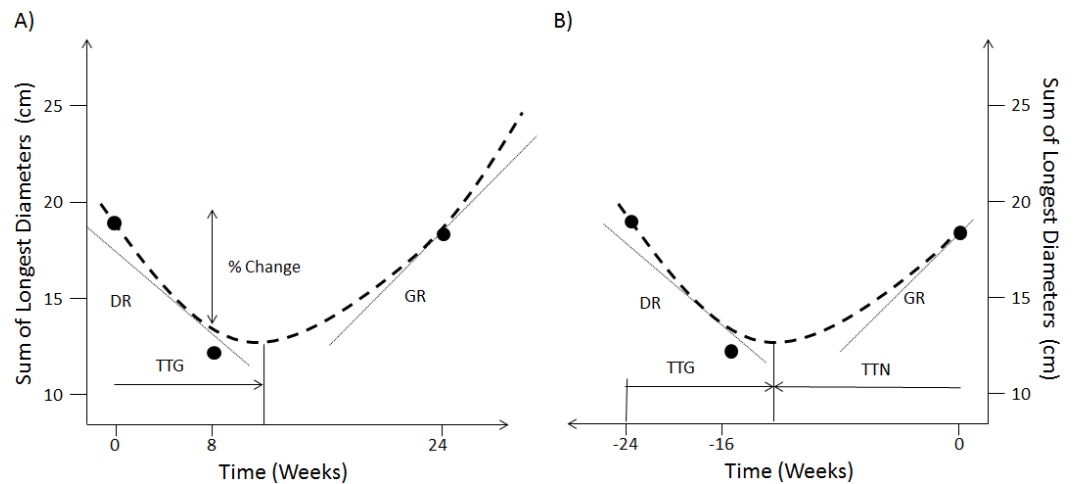


Figure 1 Forward and reverse alignment of SLD time-series. (A) Different model derived metrics, % change at week 8, TTG, DR and GR are obtained by fitting a model (dashed line) to actual observations (dots) in forward time. (B) Same model derived metrics and observations in reverse time as well as time to tumour nadir (TTN).

Full-size DOI: [10.7717/peerj.4111/fig-1](https://doi.org/10.7717/peerj.4111/fig-1)

that in the study used by Sharma et al. and/or (ii) the difference could be attributed to the choice of model based metric used: Sharma et al. used TTG whereas Kaiser chose GR.

It is important to note that within the context of this study, risk can be allocated within two different categories, individual and group risk. An individual risk metric is assessed through survival concordance probabilities (Gönen & Heller, 2005; An et al., 2011; An et al., 2015), whereas a group risk metric is assessed on whether the correct decision was made to advance a compound from phase II to phase III. In this context, individual risk estimates the patient's survival prognosis based on the patient's historical imaging time series data. Therefore, for individual risk, the time taken to acquire sufficient time-series to estimate GR and TTG must be accounted for otherwise one assumes one can see into the future. This was not done in the approaches taken in Stein et al. (2008), Stein et al. (2009), Stein et al. (2012), Blagoev et al. (2013), Claret et al. (2013) and Han et al. (2016) and so a question remains as to whether GR and TTG can be used to predict individual risk.

For group risk the question of interest is different: given historical data on a standard of care treatment in a group of patients, is a new treatment (which has mature PFS data in a group of patients) likely to produce survival improvement? This would be assessed for instance through a future prospective head-to-head trial. In this different context, full time courses can indeed be considered for metrics estimation as was the case in the approaches taken by Stein et al. (2008), Stein et al. (2009), Stein et al. (2012), Blagoev et al. (2013), Claret et al. (2013) and Han et al. (2016). It could be argued that individual risk is mainly a concern to the practicing medical community whereas group risk is a concern for the pharmaceutical industry and regulatory agencies. However, given the growth of interest in personalised (precision) medicine, individual risk is also likely to become a critical factor of future drug development.

The objective of this study was to analyse both individual and group risk using three comparator arms of phase III NSCLC (Non-Small Cell Lung Cancer) studies in the following way. For individual risk the model derived metrics, GR and TTG, calculated using the exponential decay/growth model used in previous studies ([Stein et al., 2008](#); [Stein et al., 2009](#); [Stein et al., 2012](#); [Kaiser, 2013](#); [Blagoev et al., 2013](#); [Claret et al., 2013](#); [Sharma et al., 2015](#)), were assessed via calculation of the survival concordance probability. Concordance probability was chosen for two reasons: (1) it is a measure that is routinely used when reporting the predictive capability of a prognostic model developed for use in the clinic ([Gönen & Heller, 2005](#); [An et al., 2011](#); [An et al., 2015](#)); (2) it is a non-parametric approach which I preferred over a regression model to avoid the use of a specific functional form when measuring evidence of relationship between variables. I examined the correlation to overall survival times for these quantities using two different data alignment approaches, forward and reverse. Forward alignment refers to correlating TTG and GR to overall survival without accounting for the time taken to estimate them. Reverse alignment involves re-aligning patients time-series to their time of progression and correlating TTG and GR to overall survival time minus the progression time. This approach not only accounts for the time taken to estimate GR and TTG but also assesses if those metrics are significant correlates to survival post progression. For group risk, I tested the ability of the metrics, PFS, TTG and GR, derived from the control arms of two of the phase III studies, to predict the outcome of a phase III study, where the treatments in the control arms were placed in a head-to-head trial, was tested.

METHODS

Data

Data from the control arm of three randomised phase III studies in NSCLC were collected: Erlotinib ([Scagliotti et al., 2012](#)), Docetaxel ([Ramlau et al., 2012](#)) and Paclitaxel/Carboplatin ([Socinski et al., 2012](#)). (Note, these studies were chosen as they were the only ones available at the time within ProjectDataSphere.) Both the Erlotinib and Docetaxel studies were conducted within a patient population who were pre-treated with doublet chemotherapy (majority had Paclitaxel/Carboplatin) whereas the Paclitaxel/Carboplatin study was conducted in therapy naïve patients. All three comparator arms were from studies designed to assess the overall survival of a new investigational treatment. Tumour assessments were conducted every eight weeks within the Docetaxel and Paclitaxel/Carboplatin studies and every six weeks in the Erlotinib study using RECIST 1.0. In order to calculate model derived metrics only patients that had a pre-treatment and at least one on-treatment tumour assessment were considered.

Time-series drop-out mechanism

The protocols of the studies under consideration here stated that imaging data would cease to be collected once a patient's disease had progressed. Since death is considered as a progression event I assessed what proportion of progression events were due to death. The result of this analysis determined whether or not a joint longitudinal survival analysis was performed. If the predominant drop-out mechanism was due to reasons other than death

then a joint longitudinal survival model was not considered as the drop-out mechanism is then not informative of survival.

Forward v reverse alignment

In reverse alignment the SLD measurement at the time of progression becomes time 0 and the first measurement becomes $-t$ days, see [Fig. 1B](#). Furthermore, in forwards alignment I assessed if TTG and GR correlated to overall survival. But in reverse alignment I assessed if TTG and GR correlated to overall survival minus progression free survival. Therefore in forwards alignment I shall consider TTG and GR to be biased (time-dependent bias) and in reverse alignment un-biased.

Model

In order to make the results comparable to other publications within the field ([Stein et al., 2008](#); [Stein et al., 2009](#); [Stein et al., 2012](#); [Kaiser, 2013](#); [Blagoev et al., 2013](#); [Claret et al., 2013](#); [Sharma et al., 2015](#)) I used the same decay and re-growth model as those studies to analyse the SLD time-series. The SLD is modelled as:

$$\text{SLD}(\text{time}) = A * (\exp(-B * \text{time}) + \exp(C * \text{time}) - 1).$$

The model was placed within a mixed model framework with an additive residual error model and parameters A, B and C considered to be log-normally distributed, as was done previously ([Kaiser, 2013](#); [Claret et al., 2013](#); [Sharma et al., 2015](#)). The turning point time (tp) was calculated by taking the derivative of SLD (time) with respect to time and setting it equal to 0, giving:

$$\text{tp} = (\log(B) - \log(C)) / (B + C).$$

In forward alignment this is referred to as time to tumour re-growth (TTG; [Fig. 1A](#)); in reverse alignment it is the time to nadir (TTN) from the last observation ([Fig. 1B](#)). To convert TTN to TTG, TTN is subtracted from the total imaging observation time. In forwards time B and C will be DR and GR respectively, whereas it is the other way around in reverse time. The non-linear mixed model analysis was conducted using the *nlme* package in R v3.1.1 ([Pinheiro et al., 2017](#)). Details on how the model fits to the time-series data can be found in ([Figs. S1–S6](#) contain diagnostic plots whereas [Tables S1–S6](#) contain parameter values).

Individual risk analysis

The model derived metrics GR and TTG across the three studies were assessed for their relationship to survival via an analysis of their concordance probability estimates. The concordance probability (CP) represents the probability that for any pair of patients, the patient with the better covariate value has the longer survival time. A covariate was called significant if its 95% bootstrapped (1,000 samples) concordance probability (CP) confidence intervals did not include 0.5 ([Sedgwick, 2013](#)). A value of 0.5 represents the case where there is not a consistent relationship. Concordance probability estimates were generated using the *CPE* package in R v3.1.1 ([Mo, Gonen & Heller, 2012](#)).

Table 1 Key information on the three randomised phase III studies analysed here.

	Erlotinib	Docetaxel	Paclitaxel/Carboplatin
Line of therapy	2nd	2nd	1st
Total N	369	399	413
No. death events	61	282	289
No. progression events	(301)	(353)	(328)
(Death events)	(27)	(20)	(6)
Median baseline SLD (cm)	8	8.3	10.7
25th–75th Percentile	5.2, 12.5	5.1, 12	6.9, 15.5
Median% change (weeks 6–10)	4.9	0	–16.5
25th–75th percentile	–6.6, –27.3	–10.5, –5.5	–29.8, –5.9
Median PFS (months)	3.7	4.2	6.7
(95% CI)	(2.6–3.8)	(4.1–4.7)	(6.0–7.1)
Median OS (months)	NA	10.8	12.1
(95% CI)		(9.7–12.2)	(11.2–13.2)
Median OS-PFS (months)	NA	5.2	4.1
(95% CI)		(4.3–5.8)	(3.7–5)

Notes.

SLD, Sum of Longest Diameters; PFS, Progression Free Survival; CI, Confidence Interval.

Group risk analysis

Tumour size metrics for Erlotinib and Docetaxel study arms were used to predict the outcome of a head-to-head phase III (test) trial ([Garassino et al., 2013](#)). I applied the following methodology, as used by [Sharma et al. \(2015\)](#) and [Kaiser \(2013\)](#), to generate distributions of OS HR using tumour size metrics. A total of 1,000 test trial data-sets were generated by sampling with replacement the exact same number of patients used in each study arm of the test trial ($n = 110$ for Docetaxel and $n = 112$ for Erlotinib). In generating 1,000 test trial data-sets, distributions of the following outcome measures were generated. Evaluation of TTG and GR metrics was done by calculating the ratio of the mean values between the two arms of the test trial, subsequently a distribution of these ratios were generated and for PFS the distribution of PFS HR was generated. The distribution of the predicted OS HR from TTG, GR and PFS was visualised using histograms with the actual results of the study with 95 percent confidence intervals overlaid. In addition to the visual inspection I reported the median OS HR prediction with 95 percent prediction intervals for each of the metrics.

RESULTS**Patient data and drop-out mechanism**

Details of the patients' imaging and survival characteristics can be seen in [Table 1](#). The table highlights that the number of progression events due to death were very low across all studies. Thus, the predominant reason why imaging time-series ceases to be collected is due to events other than death. The drop-out mechanism is therefore not informative of survival. Instead drop-out is informative of when the patient discontinues treatment.

Table 2 Summary of concordance probability estimates of GR (Growth Rate) and TTG (Time to Tumour Re-Growth) metrics: forward versus reverse alignments. Concordance probabilities (CP) and 95% confidence intervals (CI) are shown for GR and TTG metrics when using forwards vs. reverse alignment to analyse the time-series, for three randomised phase III studies.

	Erlotinib		Docetaxel		Docetaxel/carboplatin	
	Forward CP (95% CI)	Reverse CP (95% CI)	Forward CP (95% CI)	Reverse CP (95% CI)	Forward CP (95% CI)	Reverse CP (95% CI)
GR	0.63 (0.55–0.71)	0.49 (0.43–0.59)	0.69 (0.65–0.72)	0.52 (0.49–0.56)	0.71 (0.68–0.74)	0.53 (0.48–0.57)
TTG	0.64 (0.57–0.71)	0.53 (0.47–0.61)	0.70 (0.66–0.73)	0.54 (0.48–0.58)	0.72 (0.68–0.75)	0.55 (0.49–0.59)

Individual risk

Table 2 shows that the alignment method affects CP values. I found that GR and TTG are strong covariates for overall survival across all treatments using forward alignment; where time taken to estimate TTG and GR was not accounted for. However, I found they did not correlate to overall survival minus progression free survival, reverse alignment. These results show that GR and TTG provide no information on the survival prognosis of patients post progression.

Group risk

I then attempted to predict the outcome of the phase III test trial of Erlotinib versus Docetaxel, which showed an advantage of Docetaxel over Erlotinib, using GR, TTG or PFS. It can be seen in the histograms in Fig. 2A that the predicted distribution of OS HR using either GR (green) or PFS (grey) are close to the observed OS HR (solid vertical line) and sit well within the observed 95 percent confidence intervals (dashed vertical lines). This however is not the case for the OS HR predicted using TTG (pink) whose predicted distribution appears to sit outside the observed confidence intervals. The median and 95 percent prediction interval for the OS HR using each of the metrics, PFS, GR and TTG, were, 0.84 (0.62–1.13), 0.64 (0.53–0.77) and 0.42 (0.25–0.62) respectively. In comparing these to the actual OS HR observed, 0.73 (0.53–1), it is clear that the OS HR predicted using TTG is the furthest away. Both PFS and GR however are equally as close to the actual results with PFS under-estimating the effect and GR over-estimating the effect.

In order to investigate the difference between the predicted OS HR distributions of TTG and GR I assessed the correlation between the ratios of means between the two metrics, see Fig. 2B. The correlation plot shows that there is a relationship between the two variables ($r^2 = 0.36$) but that it is weak. This suggests that the information held within GR and TTG is moderately different and that the choice of metric could lead to different predictions. This may explain why TTG and GR did not give comparable predictions.

DISCUSSION

In this work I examined the relationship between on-treatment changes in imaging and survival outcomes, which has been gaining favour recently (Stein et al., 2008; Stein et al., 2009; Stein et al., 2012; Wang et al., 2009; Blagoev et al., 2013; Claret et al., 2013; Bruno, Mercier & Claret, 2014; Han et al., 2016). Based on the cases I have analysed, I found that:

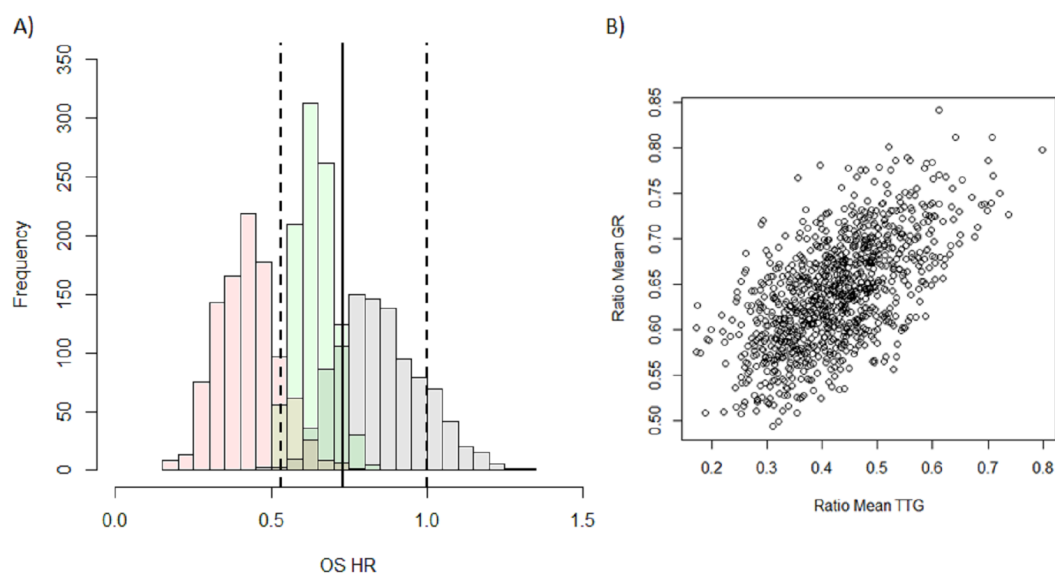


Figure 2 Distribution of model-based covariates. (A) Histograms showing the predicted OS HR distribution for PFS (grey), GR (green) and TTG (pink) obtained from the 1,000 samples of the test trial using historical data. The OS HR of the actual results of the study are overlaid: solid vertical lines is the point estimate of the HR with dashed lines being the 95 percent confidence intervals. (B) Correlation between the ratio of mean TTG versus ratio of mean GR between each arm for each test trial.

Full-size  DOI: [10.7717/peerj.4111/fig-2](https://doi.org/10.7717/peerj.4111/fig-2)

- (1) Model derived metrics TTG and GR were not significant covariates of survival post progression.
- (2) The outcome of the phase III study using historical data was predicted just as well by conventional PFS HR as TTG and GR.

Regarding the first point, model derived metrics TTG and GR were found to be significant correlates for overall survival when the time taken to estimate them was not accounted for i.e., they were considered to be biased. However, by re-phrasing the question to, do TTG and GR correlate to survival post progression? I accounted for this bias, since imaging data ceases to be collected once a patients disease progresses. The result of re-phrasing the question highlighted that the dynamics of the disease leading up to progression have no bearing on the prognosis of a patient post progression. This result may seem surprising however it may not be once one considers what happens at the time of progression. The patient stops taking the current line of treatment and at some time-point post-progression may start a new treatment. Furthermore, this new treatment may be quite variable if there is no standard treatment approved. Therefore unless there is a correlation between the dynamics leading up to progression and the dynamics under the new unknown treatment one would expect the results found here to hold in most clinical trials.

Regarding the second finding, I assessed three metrics in their ability to correctly predict the outcome of a phase III trial, an assessment of group risk, which was a head-to-head test of two of the drugs used in this analysis, Erlotinib and Docetaxel. The three metrics assessed were PFS HR, TTG and GR. Both PFS and GR were as far away to the actual result as each other; with GR over-predicting and PFS under-predicting the advantage of Docetaxel over

Erlotinib. TTG, however, was furthest away from the actual result; it overestimated the advantage of Docetaxel over Erlotinib more so than GR. These results confirm previous findings where PFS was equal or superior to GR in making the correct decision in moving from phase II to phase III ([Kaiser, 2013](#)). This is in contrast with the study by [Sharma et al. \(2015\)](#) where PFS was found to be slightly less predictive than TTG. A reason for this could be that the study used here as the test case was powered to show a PFS difference which was not the case in the study done by Sharma et al., where they stopped the re-sampled study after 18 weeks and so was not powered to detect a PFS difference which is usually done for phase II studies. Sharma et al. proposed that the difference in magnitude of PFS between their study arms and that of Kaiser explained why the latter found PFS superior to GR. However, the PFS HR I investigated here (0.71) was similar to the one in the study by Sharma et al. (HR: 0.79). I suggest that the reason for the difference in results between Sharma et al. and Kaiser is related to the choice of model derived metric. Sharma et al. used TTG whereas Kaiser et al. used GR. The results here show that the ratio of means of those metrics between treatment arms do not correlate strongly. Therefore, the discrepancy could be due to choice of metric and not to differences in PFS. It is interesting to note that although TTG and GR were not strong covariates for survival post-progression, they performed well at predicting group risk. Thus, it is possible that a different approach should be used for modelling individual and group risk.

Although the analyses performed were for only one disease type, NSCLC, and were within a limited pharmacological space, they highlight the importance of accounting for the time it takes to estimate model derived metrics when assessing their prognostic value for individual risk. In addition to the approach described here regarding re-aligning a patients time-series to their time of progression an alternative approach would be to perform a landmark analysis in forward alignment at a specific time-point as was done by [Sharma et al. \(2015\)](#). Both of these approaches account for the time taken to estimate TTG and GR. For group risk, although the results here show that PFS HR can predict OS HR, it has been shown that in general this may not be the case ([Blumenthal et al., 2015](#)). However, this does not suggest that PFS cannot be used for decision making as I found that for majority of patients radiological progression precedes death; it merely implies that it may not always predict quantitatively what the OS difference could be.

Overall, when comparing the results seen here with previous published studies assessing the decision making value of tumour size metrics, it suggests that PFS remains the best metric to use for decision making when considering group risk. For individual risk, model derived metrics derived from SLD time-series might be inappropriate. This is in contrast to prognostic models developed using standard pre-treatment clinical variables (e.g., lactate dehydrogenase, alkaline phosphatase etc.) which have been shown to be predictive for a specific drug within a specific disease setting ([Halabi et al., 2014](#); [Wendling et al., 2016](#)). It would be interesting to further investigate how prognostic models developed using routine clinical variables could be adapted to be used for predicting group risk and how well these perform compared to PFS for decision making and quantitatively predicting treatment differences in OS.

ACKNOWLEDGEMENTS

The author wishes to thank Professor Leon Aarons, Dr. Fernando Ortega, Dr. Giovanni Di Veroli and Dr James Yates for helpful discussions. This publication is based on research using information obtained from <http://www.projectdatasphere.org>, which is maintained by Project Data Sphere, LLC. Neither Project Data Sphere, LLC nor the owner(s) of any information from the web site have contributed to, approved or are in any way responsible for the contents of this publication.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

Hitesh B. Mistry was funded by Manchester School of Pharmacy through an AstraZeneca grant. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the author:
Manchester School of Pharmacy.

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Hitesh B. Mistry conceived and designed the experiments, performed the experiments, analyzed the data, contributed reagents/materials/analysis tools, wrote the paper, prepared figures and/or tables, reviewed drafts of the paper.

Data Availability

The following information was supplied regarding data availability:

The data is available through the ProjectDataSphere platform, which doesn't allow re-publication of their data.

<https://www.projectdatasphere.org/projectdatasphere/html/home>.

All users have access to the data after registration: The 3 data-sets of interest used were:

<https://www.projectdatasphere.org/projectdatasphere/html/content/133>;

<https://www.projectdatasphere.org/projectdatasphere/html/content/115>;

<https://www.projectdatasphere.org/projectdatasphere/html/content/108>.

Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.4111#supplemental-information>.

REFERENCES

- An M-W, Dong X, Meyers J, Han Y, Grothey A, Bogaerts J, Sargent DJ, Mandrekar SJ. 2015. Evaluating continuous tumor measurement-based metrics as phase II endpoints for predicting overall survival. *Journal of the National Cancer Institute* 107(11):djv239 DOI 10.1093/jnci/djv239.
- An M-W, Mandrekar SJ, Branda ME, Hillman SL, Adjei AA, Pitot HC, Goldberg RM, Sargent DJ. 2011. Comparison of continuous versus categorical tumor measurement-based metrics to predict overall survival in cancer treatment trials. *Clinical Cancer Research* 17:6592–6599 DOI 10.1158/1078-0432.CCR-11-0822.
- Blagoev KB, Wilkerson J, Stein WD, Motzer RJ, Bates SE, Fojo AT. 2013. Sunitinib does not accelerate tumor growth in patients with metastatic renal cell carcinoma. *Cell Reports* 3:277–281 DOI 10.1016/j.celrep.2013.01.015.
- Blumenthal GM, Karuri SW, Zhang H, Zhang L, Khozin S, Kazandjian D, Tang S, Sridhara R, Keegan P, Pazdur R. 2015. Overall response rate, progression-free survival, and overall survival with targeted and standard therapies in advanced non-small-cell lung cancer: US Food and Drug Administration trial-level and patient-level analyses. *Journal of Clinical Oncology* 33:1008–1014 DOI 10.1200/JCO.2014.59.0489.
- Bruno R, Mercier F, Claret L. 2014. Evaluation of tumor size response metrics to predict survival in oncology clinical trials. *Clinical Pharmacology and Therapeutics* 95:386–393 DOI 10.1038/clpt.2014.4.
- Claret L, Gupta M, Han K, Joshi A, Sarapa N, He J, Powell B, Bruno R. 2013. Evaluation of tumor-size response metrics to predict overall survival in western and chinese patients with first-line metastatic colorectal cancer. *Journal of Clinical Oncology* 17:2110–2114 DOI 10.1200/JCO.2012.45.0973.
- Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, Dancey J, Arbuck S, Gwyther S, Mooney M, Rubinstein L, Shankar L, Dodd L, Kaplan R, Lacombe D, Verweij J. 2009. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *European Journal of Cancer* 45:228–247 DOI 10.1016/j.ejca.2008.10.026.
- Garassino MC, Martelli O, Broggin M, Farina G, Veronese S, Rulli E, Bianchi F, Bettini A, Longo F, Moscetti L, Tomirotti M, Marabese M, Ganzinelli M, Lauricella C, Labianca R, Floriani I, Giaccone G, Torri V, Scanni A, Marsoni S. 2013. Erlotinib versus docetaxel as second-line treatment of patients with advanced non-small-cell lung cancer and wild-type EGFR tumours (TAILOR): a randomised controlled trial. *The Lancet Oncology* 14:981–988 DOI 10.1016/S1470-2045(13)70310-3.
- Gönen M, Heller G. 2005. Concordance probability and discriminatory power in proportional hazards regression. *Biometrika* 92:965–970 DOI 10.1093/biomet/92.4.965.
- Halabi S, Lin C-Y, Kelly WK, Fizazi KS, Moul JW, Kaplan EB, Morris MJ, Small EJ. 2014. Updated prognostic model for predicting overall survival in first-line chemotherapy for patients with metastatic castration-resistant prostate cancer. *Journal of Clinical Oncology* 7:671–677 DOI 10.1200/JCO.2013.52.3696.

- Han K, Claret L, Piao Y, Hegde P, Joshi A, Powell J, Jin J, Bruno R. 2016.** Simulations to predict clinical trial outcome of bevacizumab plus chemotherapy vs., chemotherapy alone in patients with first-line gastric cancer and elevated plasma VEGF-A. *CPT: Pharmacometrics & Systems Pharmacology* 5:352–358 DOI 10.1002/psp4.12064.
- Ibrahim JG, Chu H, Chen LM. 2010.** Basic concepts and methods for joint models of longitudinal and survival data. *Journal of Clinical Oncology* 28:2796–2801 DOI 10.1200/JCO.2009.25.0654.
- Kaiser LD. 2013.** Tumor burden modeling versus progression-free survival for phase II decision making. *Clinical Cancer Research* 19:314–319 DOI 10.1158/1078-0432.CCR-12-2161.
- Mansmann UR, Laubender RP. 2013.** Methodologic diligence is needed to define and validate tumor-size response metrics to predict overall survival in first-line metastatic colorectal cancer. *Journal of Clinical Oncology* 31:4373–4374 DOI 10.1200/JCO.2013.51.2954.
- Mistry H. 2016.** Time-dependent bias of tumor growth rate and time to tumor regrowth. *CPT: Pharmacometrics & Systems Pharmacology* 5:587 DOI 10.1002/psp4.12145.
- Mo Q, Gonen M, Heller G. 2012.** CPE: concordance probability estimates in survival analysis. Available at <https://CRAN.R-project.org/package=CPE>.
- Pinheiro J, Bates D, DebRoy S, Sarkar D, R Core Team. 2017.** nlme: linear and nonlinear mixed effects models. R package version 3.1.1. Available at <https://cran.r-project.org/web/packages/nlme/index.html>.
- Ramlau R, Gorbunova V, Ciuleanu TE, Novello S, Ozguroglu M, Goksel T, Baldotto C, Bennouna J, Shepherd FA, Le-Guenec S, Rey A, Miller V, Thatcher N, Scagliotti G. 2012.** Afibercept and docetaxel versus docetaxel alone after platinum failure in patients with advanced or metastatic non-small-cell lung cancer: a randomized, controlled phase III Trial. *Journal of Clinical Oncology* 30:3640–3647 DOI 10.1200/JCO.2012.42.6932.
- Scagliotti GV, Krzakowski M, Szczesna A, Strausz J, Makhson A, Reck M, Wierzbiicki RF, Albert I, Thomas M, Miziara JEA, Papai ZS, Karaseva N, Thongprasert S, Portulas ED, Von Pawel J, Zhang K, Selaru P, Tye L, Chao RC, Govindan R. 2012.** Sunitinib plus erlotinib versus placebo plus erlotinib in patients with previously treated advanced non-small-cell lung cancer: a phase III trial. *Journal of Clinical Oncology* 17:2070–2078 DOI 10.1200/JCO.2011.39.2993.
- Sedgwick P. 2013.** P values or confidence intervals? *BMJ* 346:f3212 DOI 10.1136/bmj.f3212.
- Sharma MR, Gray E, Goldberg RM, Sargent DJ, Karrison TG. 2015.** Resampling the N9741 trial to compare tumor dynamic versus conventional end points in randomized phase II trials. *Journal of Clinical Oncology* 33:36–41 DOI 10.1200/JCO.2014.57.2826.
- Socinski MA, Bondarenko I, Karaseva NA, Makhson AM, Vynnychenko I, Okamoto I, Hon JK, Hirsh V, Bhar P, Zhang H, Iglesias JL, Renschler MF. 2012.** Weekly nab-paclitaxel in combination with carboplatin versus solvent-based paclitaxel plus carboplatin as first-line therapy in patients with advanced non-small-cell lung

- cancer: final results of a phase III trial. *Journal of Clinical Oncology* **30**:2055–2062 DOI [10.1200/JCO.2011.39.5848](https://doi.org/10.1200/JCO.2011.39.5848).
- Stein WD, Huang H, Menefee M, Edgerly M, Kotz H, Dwyer A, Yang J, Bates SE. 2009.** Other paradigms: growth rate constants and tumor burden determined using computed tomography data correlate strongly with the overall survival of patients with renal cell carcinoma. *Cancer Journal* **15**:441–447 DOI [10.1097/PPO.0b013e3181be1b90](https://doi.org/10.1097/PPO.0b013e3181be1b90).
- Stein WD, Wilkerson J, Kim ST, Huang X, Motzer RJ, Fojo AT, Bates SE. 2012.** Analyzing the pivotal trial that compared sunitinib and interferon alfa in renal cell carcinoma, using a method that assesses tumor regression and growth. *Clinical Cancer Research* **18**:2374–2381 DOI [10.1158/1078-0432.CCR-11-2275](https://doi.org/10.1158/1078-0432.CCR-11-2275).
- Stein WD, Yang J, Bates SE, Fojo T. 2008.** Bevacizumab reduces the growth rate constants of renal carcinomas: a novel algorithm suggests early discontinuation of bevacizumab resulted in a lack of survival advantage. *The Oncologist* **13**:1055–1062 DOI [10.1634/theoncologist.2008-0016](https://doi.org/10.1634/theoncologist.2008-0016).
- Therasse P, Arbuck SG, Eisenhauer EA, Wanders J, Kaplan RS, Rubinstein L, Verweij J, Van Glabbeke M, Van Oosterom AT, Christian MC, Gwyther SG. 2000.** New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada. *Journal of the National Cancer Institute* **92**:205–216 DOI [10.1093/jnci/92.3.205](https://doi.org/10.1093/jnci/92.3.205).
- Toffart A-C, Moro-Sibilot D, Couraud S, Merle P, Perol M, Girard N, Souquet P-J, Mastroianni B, Ferretti GR, Romand P, Chatellain P, Vesin A, Brambilla E, Brambilla C, Timsit J-F. 2014.** Evaluation of RECIST in chemotherapy-treated lung cancer: the pharmacogenoscan study. *BMC Cancer* **14**:989 DOI [10.1186/1471-2407-14-989](https://doi.org/10.1186/1471-2407-14-989).
- Van Walraven C, Davis D, Forster AJ, Wells GA. 2004.** Time-dependent bias was common in survival analyses published in leading clinical journals. *Journal of Clinical Epidemiology* **57**:672–682 DOI [10.1016/j.jclinepi.2003.12.008](https://doi.org/10.1016/j.jclinepi.2003.12.008).
- Wang Y, Sung C, Dartois C, Ramchandani R, Booth BP, Rock E, Gobburu J. 2009.** Elucidation of relationship between tumor size and survival in non-small-cell lung cancer patients can aid early decision making in clinical drug development. *Clinical Pharmacology and Therapeutics* **86**:167–174 DOI [10.1038/clpt.2009.64](https://doi.org/10.1038/clpt.2009.64).
- Wendling T, Mistry H, Ogungbenro K, Aarons L. 2016.** Predicting survival of pancreatic cancer patients treated with gemcitabine using longitudinal tumour size data. *Cancer Chemotherapy and Pharmacology* **77**:927–938 DOI [10.1007/s00280-016-2994-x](https://doi.org/10.1007/s00280-016-2994-x).