bioRxiv preprint doi: https://doi.org/10.1101/109934; this version posted February 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

A cautionary tale on using tumour growth rate to predict survival

Hitesh B. Mistry[1*] and Fernando Ortega[1]

[1]Centre For Drug Disease Modelling & Simulation, Manchester Pharmacy School, University of Manchester

*corresponding author

Corresponding email: hitesh.mistry@manchester.ac.uk

Corresponding address: Centre For Drug Disease Modelling & Simulation, Manchester Pharmacy School, University of Manchester

bioRxiv preprint doi: https://doi.org/10.1101/109934; this version posted February 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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. We 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 we 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.

Keywords: Biostatistics, Longitudinal Imaging, Survival

bioRxiv preprint doi: https://doi.org/10.1101/109934; this version posted February 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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) (1,2). RECIST v1.0 and v1.1 involves the evaluation of a drugs' efficacy against both target and non-target lesions. The drugs 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 us 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 we assume 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

bioRxiv preprint doi: https://doi.org/10.1101/109934; this version posted February 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 (3–5). 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 (6–13). **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 (7–10,12,13) 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 (14,15). 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 (16,17).

Contradictory results have been reported regarding the predictive power of these metrics. In one hand, Sharma el al. (18) 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 (19) used a larger study and found that PFS was a superior endpoint compared to GR in 5 of his 6 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 that in the study used by Sharma et al. and/or ii) the difference could be attributed to the choice of model

bioRxiv preprint doi: https://doi.org/10.1101/109934; this version posted February 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

based metric used: Sharma et al. used TTG whereas Kaiser chose GR.

It is important to note that in our context, risk can be allocated within two different categories, individual and group risk. An individual risk metric is assessed through survival concordance probabilities (3-5,20), 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 we assume we can see into the future. This was not done in the approaches taken in (7–10,12,13) 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 (7–10,12,13). 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 (7–10,12,13,18,19), were assessed via calculation of the survival concordance probability. Concordance probability was chosen

bioRxiv preprint doi: https://doi.org/10.1101/109934; this version posted February 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 (3-5,20); 2) it is a non-parametric approach which we preferred over a regression model to avoid the use of a specific functional form when measuring evidence of relationship between variables. We 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, we 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 (21), Docetaxel (22) and Paclitaxel/Carboplatin (23). Tumour assessments were conducted every 8 weeks within the Docetaxel and Paclitaxel/Carboplatin studies and every 6 weeks in the Erlotinib study using RECIST 1.0. In order to compare model derived metrics to empirical early changes in tumour size and RECIST response categorisation only patients that had a pre-treatment and at least one on-treatment tumour assessment were considered.

bioRxiv preprint doi: https://doi.org/10.1101/109934; this version posted February 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 we 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

Forward and reverse alignment involved aligning patients' tumour time-series from the time of randomisation or from the time of progression respectively. Therefore in reverse alignment the SLD measurement at the time of progression becomes time 0 and the first measurement becomes -*t* days, see **Figure 1B**. Furthermore, in forwards alignment we assessed if TTG and GR correlated to overall survival. But in reverse alignment we assessed if TTG and GR correlated to overall survival minus progression free survival. Therefore in forwards alignment we shall consider TTG and GR to be biased (time-dependent bias) and in reverse alignment un-biased.

Model

In order to make our results comparable to other publications within the field (7–10,12,13,18,19) we used the same decay and re-growth model as those studies to analyse the SLD time-series. The SLD is modelled as:

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

bioRxiv preprint doi: https://doi.org/10.1101/109934; this version posted February 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 per. 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:

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

In forward alignment this is referred to as time to tumour re-growth (TTG; **Figure 1A**); in reverse alignment it is the time to nadir (TTN) from the last observation (**Figure 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. Details on the model fits to the time-series data can be found in supplementary information (Figures 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 (1000 samples) concordance probability (CP) confidence intervals did not include 0.5 (24). 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.

Group Risk Analysis

bioRxiv preprint doi: https://doi.org/10.1101/109934; this version posted February 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Tumour size metrics for Erlotinib and Docetaxel study arms were used to predict the outcome of a head-to-head phase III (test) trial. We applied the following methodology, as used by Sharma et al. (18) and Kaiser (19), to generate distributions of OS HR using tumour size metrics. A total of 1000 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 1000 test trial data-sets, distributions of the following outcome measures were generated. Evaluation of TTG and GR metrics, calculated via both forward (FA) and reverse alignment (RA), 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.

Research Ethics

No ethical approval was required for this analysis as the data was obtained via an open data portal where the patients were de-identified: www.projectdatasphere.org. No animals were used in this study and nor was any fieldwork conducted.

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.

Individual Risk

Table 2 shows that the alignment method affects CP values. We found that GR and TTG are strong covariates for overall survival across all treatments using forward alignment;

bioRxiv preprint doi: https://doi.org/10.1101/109934; this version posted February 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

where time taken to estimate TTG and GR was not accounted for. However, we 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 prognosis of patients post progression.

Group Risk

We then attempted to predict the outcome of the phase III test trial of Erlotinib versus Docetaxel using GR, TTG or PFS. GR and TTG were calculated using both forwards and reverse alignment methods. It can be seen in **Figure 2A** that of the metrics analysed, all except TTG, calculated using forwards alignment, performed equally well at predicting the OS HR. **Figure 2B** shows that there is no strong correlation between the ratio of mean TTG and ratio of mean GR in either forwards or reverse alignment. This observation suggests that the information held within GR and TTG is different and that the choice of metric can lead to different predictions. This may explain why TTG and GR estimated using forward alignment did not give comparable predictions.

Discussion

In this work we examined the relationship between on-treatment changes in imaging and survival outcomes, which has been gaining favour recently (6–13). Based on the cases we have analysed, we found that:

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.

bioRxiv preprint doi: https://doi.org/10.1101/109934; this version posted February 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Regarding our 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, we 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.

Regarding our second finding, we assessed five 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 five metrics assessed were PFS HR, TTG and GR. (Both forward and reverse alignment was used to calculate TTG and GR.) Most metrics agreed with the results of the actual phase III trial, with PFS being as good as predictor as the best model derived one. 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 (19). This is in contrast with the study by Sharma et al (18) 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 we investigated here (0.71) was similar to the one in the study by Sharma et al. (HR: 0.79). We 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

bioRxiv preprint doi: https://doi.org/10.1101/109934; this version posted February 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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. 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 (25). However, this does not suggest that PFS cannot be used for decision making as we found that for most 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 (26,27). 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.

bioRxiv preprint doi: https://doi.org/10.1101/109934; this version posted February 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Data Availability

All data used in this analysis is available freely through the ProjectDataSphere portal: www.projectdatasphere.org.

Competing Interests

I/we have no competing interests.

Authors' Contributions

HBM conceived the study, extracted the data, participated in data analysis, carried out the statistical analyses, participated in the design of the study and drafted the manuscript; FO c coordinated the study and helped draft the manuscript. All authors gave final approval for publication.

Funding

Hitesh B. Mistry and Fernando Ortega were funded by Manchester School of Pharmacy through an AstraZeneca grant.

Acknowledgements

The authors wish to thank Professor Leon Aarons, Dr. Giovanni Di Veroli, Dr James Yates and Dr. Andrew Stein for helpful discussions. This publication is based on research using information obtained from 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.

bioRxiv preprint doi: https://doi.org/10.1101/109934; this version posted February 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

1. Therasse P, Arbuck SG, Eisenhauer EA, Wanders J, Kaplan RS, Rubinstein L, et al. 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. J Natl Cancer Inst. 2000 Feb 2;92(3):205–16.

2. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). Eur J Cancer Oxf Engl 1990. 2009 Jan;45(2):228–47.

3. An M-W, Dong X, Meyers J, Han Y, Grothey A, Bogaerts J, et al. Evaluating Continuous Tumor Measurement-Based Metrics as Phase II Endpoints for Predicting Overall Survival. J Natl Cancer Inst. 2015 Nov;107(11).

4. An M-W, Mandrekar SJ, Branda ME, Hillman SL, Adjei AA, Pitot HC, et al. Comparison of Continuous versus Categorical Tumor Measurement–Based Metrics to Predict Overall Survival in Cancer Treatment Trials. Clin Cancer Res. 2011 Oct 15;17(20):6592–9.

5. Toffart A-C, Moro-Sibilot D, Couraud S, Merle P, Perol M, Girard N, et al. Evaluation of RECIST in chemotherapy-treated lung cancer: the Pharmacogenoscan Study. BMC Cancer. 2014 Dec 20;14(1):989.

bioRxiv preprint doi: https://doi.org/10.1101/109934; this version posted February 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

6. Bruno R, Mercier F, Claret L. Evaluation of tumor size response metrics to predict survival in oncology clinical trials. Clin Pharmacol Ther. 2014 Apr;95(4):386–93.

7. Claret L, Gupta M, Han K, Joshi A, Sarapa N, He J, et al. Evaluation of Tumor-Size Response Metrics to Predict Overall Survival in Western and Chinese Patients With First-Line Metastatic Colorectal Cancer. J Clin Oncol. 2013 May 6;JCO.2012.45.0973.

8. Stein WD, Huang H, Menefee M, Edgerly M, Kotz H, Dwyer A, et al. 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 J Sudbury Mass. 2009;15(5):441–7.

9. Stein WD, Wilkerson J, Kim ST, Huang X, Motzer RJ, Fojo AT, et al. Analyzing the pivotal trial that compared sunitinib and interferon alfa in renal cell carcinoma, using a method that assesses tumor regression and growth. Clin Cancer Res. 2012 Apr 15;18(8):2374–81.

10. Stein WD, Yang J, Bates SE, Fojo T. 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. 2008 Oct;13(10):1055–62.

11. Wang Y, Sung C, Dartois C, Ramchandani R, Booth BP, Rock E, et al. Elucidation of relationship between tumor size and survival in non-small-cell lung cancer patients can aid early decision making in clinical drug development. Clin Pharmacol Ther. 2009 Aug;86(2):167–74.

bioRxiv preprint doi: https://doi.org/10.1101/109934; this version posted February 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

12. Blagoev KB, Wilkerson J, Stein WD, Motzer RJ, Bates SE, Fojo AT. Sunitinib Does Not Accelerate Tumor Growth in Patients with Metastatic Renal Cell Carcinoma. Cell Rep. 2013 Feb 21;3(2):277–81.

13. Han K, Claret L, Piao Y, Hegde P, Joshi A, Powell J, et al. 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 Pharmacomet Syst Pharmacol. 2016 Jul 1;5(7):352–8.

14. Ibrahim JG, Chu H, Chen LM. Basic Concepts and Methods for Joint Models of Longitudinal and Survival Data. J Clin Oncol. 2010 Jun 1;28(16):2796–801.

15. Mansmann UR, Laubender RP. Methodologic Diligence Is Needed to Define and Validate Tumor-Size Response Metrics to Predict Overall Survival in First-Line Metastatic Colorectal Cancer. J Clin Oncol. 2013 Dec 1;31(34):4373–4.

16. van Walraven C, Davis D, Forster AJ, Wells GA. Time-dependent bias was common in survival analyses published in leading clinical journals. J Clin Epidemiol. 2004 Jul;57(7):672–82.

17. Mistry H. Time-Dependent Bias of Tumor Growth Rate and Time to Tumor Regrowth. CPT Pharmacomet Syst Pharmacol. 2016 Nov 1;5(11):587–587.

18. Sharma MR, Gray E, Goldberg RM, Sargent DJ, Karrison TG. Resampling the N9741 Trial to Compare Tumor Dynamic Versus Conventional End Points in Randomized Phase II Trials. J Clin Oncol. 2015 Jan 1;33(1):36–41.

19. Kaiser LD. Tumor Burden Modeling Versus Progression-Free Survival for Phase II Decision Making. Clin Cancer Res. 2013 Jan 15;19(2):314–9.

bioRxiv preprint doi: https://doi.org/10.1101/109934; this version posted February 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

20. Gönen M, Heller G. Concordance probability and discriminatory power in proportional hazards regression. Biometrika. 2005 Dec 1;92(4):965–70.

21. Scagliotti GV, Krzakowski M, Szczesna A, Strausz J, Makhson A, Reck M, et al. Sunitinib Plus Erlotinib Versus Placebo Plus Erlotinib in Patients With Previously Treated Advanced Non–Small-Cell Lung Cancer: A Phase III Trial. J Clin Oncol. 2012 May 7;JCO.2011.39.2993.

22. Ramlau R, Gorbunova V, Ciuleanu TE, Novello S, Ozguroglu M, Goksel T, et al. Aflibercept 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. J Clin Oncol. 2012 Oct 10;30(29):3640–7.

23. Socinski MA, Smit EF, Lorigan P, Konduri K, Reck M, Szczesna A, et al. Phase III study of pemetrexed plus carboplatin compared with etoposide plus carboplatin in chemotherapy-naive patients with extensive-stage small-cell lung cancer. J Clin Oncol Off J Am Soc Clin Oncol. 2009 Oct 1;27(28):4787–92.

24. Sedgwick P. P values or confidence intervals? BMJ. 2013 May 17;346:f3212.

25. Blumenthal GM, Karuri SW, Zhang H, Zhang L, Khozin S, Kazandjian D, et al. 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. J Clin Oncol. 2015 Mar 20;33(9):1008–14.

26. Halabi S, Lin C-Y, Kelly WK, Fizazi KS, Moul JW, Kaplan EB, et al. Updated Prognostic Model for Predicting Overall Survival in First-Line Chemotherapy for

bioRxiv preprint doi: https://doi.org/10.1101/109934; this version posted February 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Patients With Metastatic Castration-Resistant Prostate Cancer. J Clin Oncol. 2014 Jan 21;JCO.2013.52.3696.

27. Wendling T, Mistry H, Ogungbenro K, Aarons L. Predicting survival of pancreatic cancer patients treated with gemcitabine using longitudinal tumour size data. Cancer Chemother Pharmacol. 2016;77(5):927–38.

bioRxiv preprint doi: https://doi.org/10.1101/109934; this version posted February 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figures

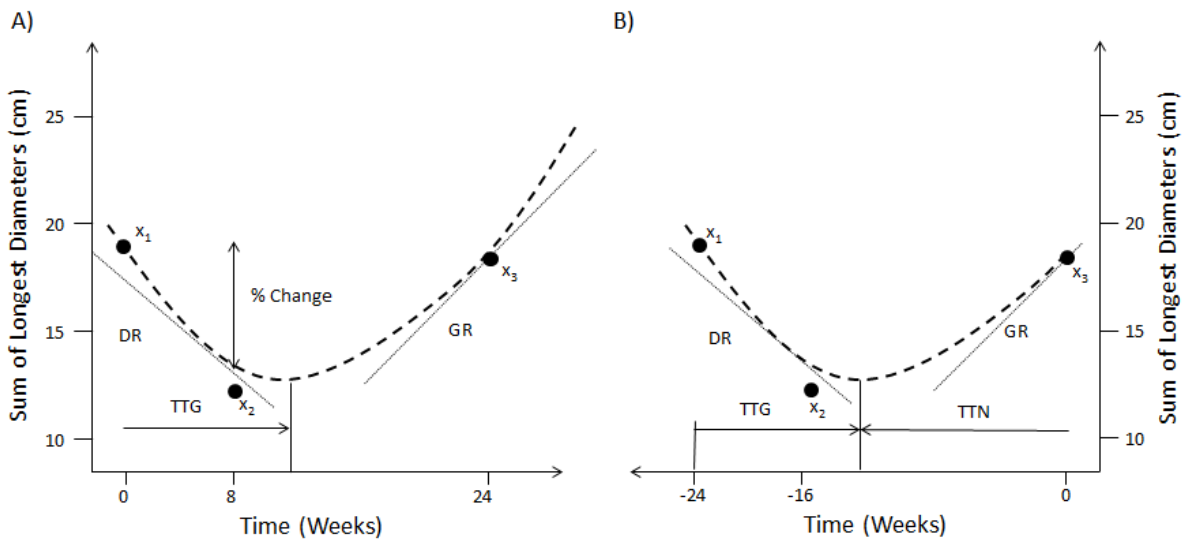

Figure 1 Forward and reverse alignment of SLD time-series A) Different model derived metrics, % change at week 8, TTG, DR and GR are obtained by fitting a model (dashed line) to actual observations (dots) in forward time. B) Same model derived metrics and observations in reverse time as well as time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

bioRxiv preprint doi: https://doi.org/10.1101/109934; this version posted February 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

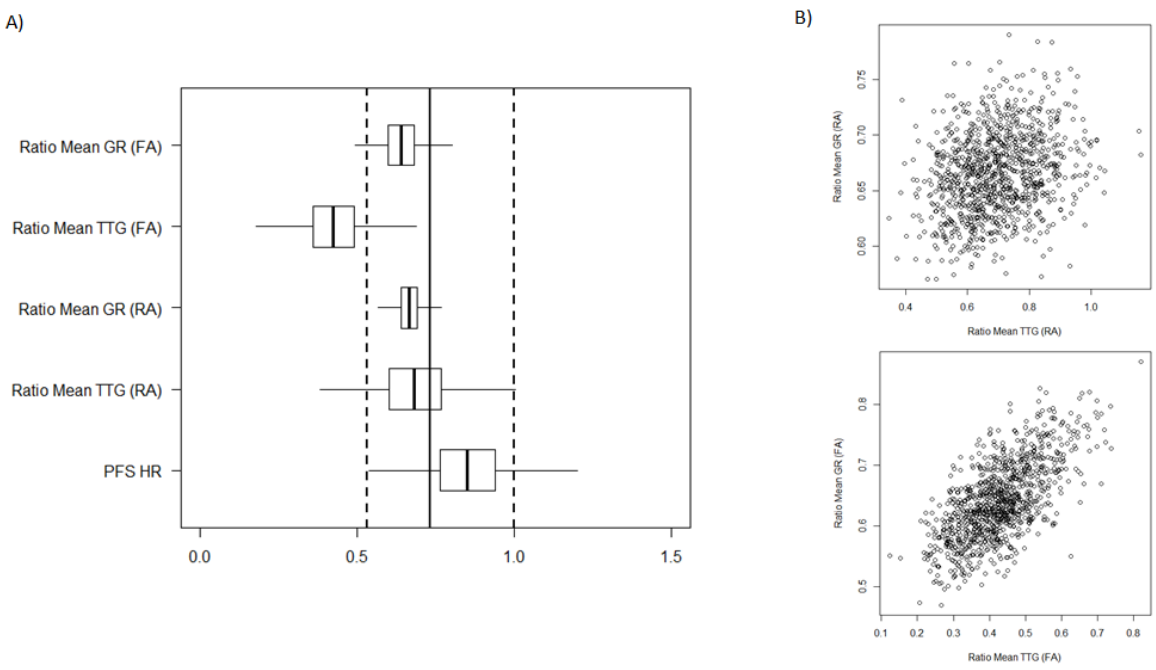

Figure 2 Distribution of model-based covariates, **A)** Boxplots show the distribution of measures obtained from the 1000 samples of the test trial using historical data. The OS HR of the actual results of the study are overlaid: solid vertical lines is the point estimate of the HR with dashed lines being the 95 % confidence intervals. **B)** Correlation between the ratio of mean TTG versus ratio of mean GR between each arm for each test trial for FA and reverse RA alignment.

bioRxiv preprint doi: https://doi.org/10.1101/109934; this version posted February 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Tables

	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)

Table 1 Key information for three randomised phase III studies analysed here. (SLD: Sum of Longest Diameters; PFS: Progression Free Survival; C.I.: Confidence Interval)

bioRxiv preprint doi: https://doi.org/10.1101/109934; this version posted February 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

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 (C.I.) are shown for GR and TTG metrics when using forwards vs. reverse alignment to analyse the time-series, for three randomised phase III studies.

bioRxiv preprint doi: https://doi.org/10.1101/109934; this version posted February 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplementary Information

This document contains the diagnostic plots and tables of parameter values for the SLD time-series analysis conducted using the model described in the main paper.

Erlotinib – Forwards Alignment

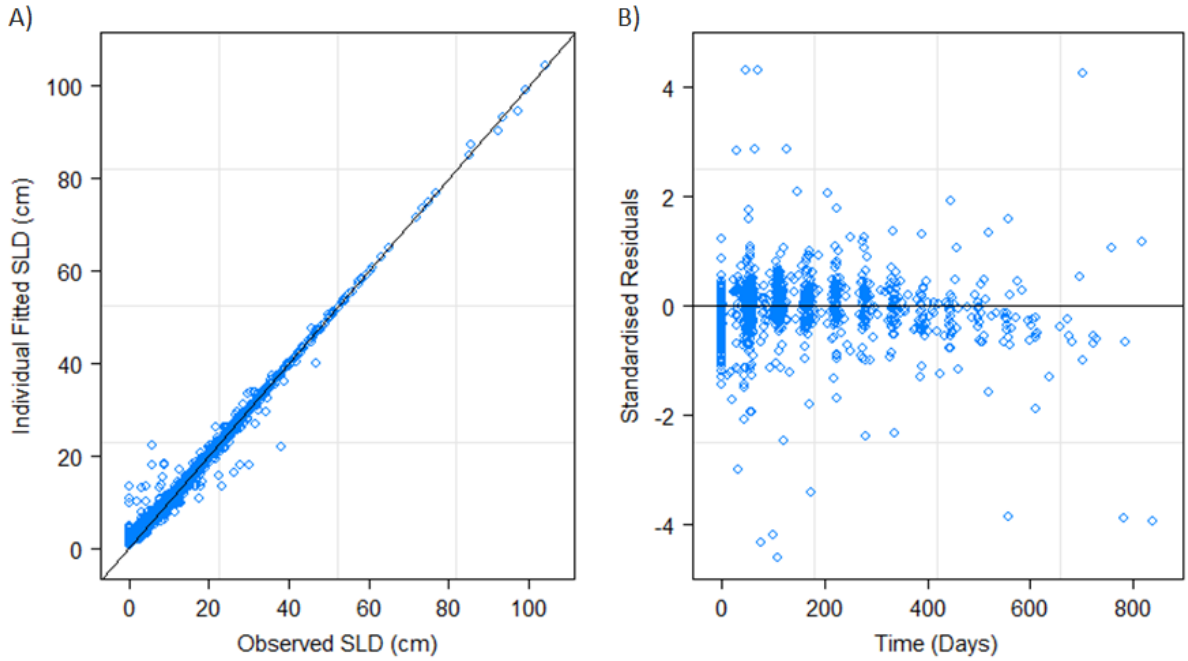

Figure 1 Panel A shows the observed SLD values against the individual fitted. Panel B shows the individual standardised residuals over time.

bioRxiv preprint doi: https://doi.org/10.1101/109934; this version posted February 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

log(A)		log(B)		log(C)		Residual
Mean (95% C.I.)	S.D. (95% C.I.)	Mean (95% C.I.)	S.D. (95% C.I.)	Mean (95% C.I.)	S.D. (95% C.I.)	S.E. (95% C.I.)
2.79 (2.72, 2.85)	0.62 (0.57, 0.67)	-5.04 (-5.22, -4.85)	1.09 (0.94, 1.26)	-5.60 (-5.72, -5.48)	0.89 (0.80, 0.97)	2.35 (2.22, 2.49)

Table 1 Provides estimated mean and standard deviation (S.D.) values together with approximate 95% confidence intervals for the parameter distributions of the model described in the main paper. Estimate of the within group standard error (S.E.) of the residuals with 95% CI is also provided.

Erlotinib – Reverse Alignment

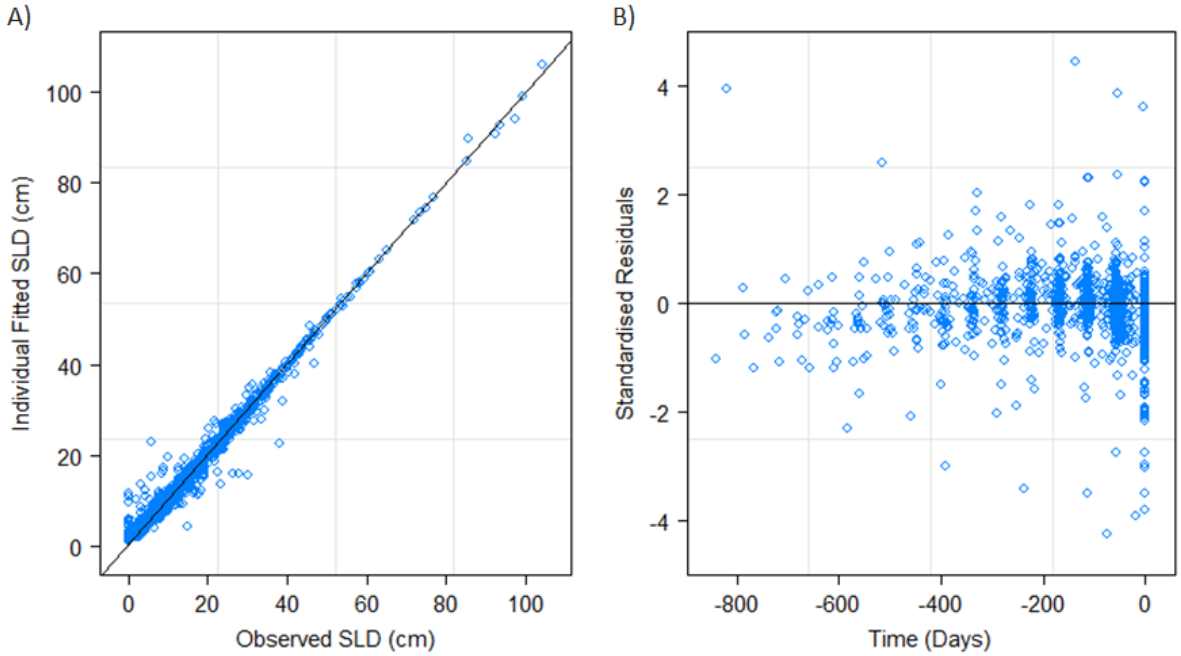

Figure 2 Panel A shows the observed SLD values against the individual fitted. Panel B shows the individual standardised residuals over time.

bioRxiv preprint doi: https://doi.org/10.1101/109934; this version posted February 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

log(A)		log(B)		log(C)		Residual
Mean (95% C.I.)	S.D. (95% C.I.)	Mean (95% C.I.)	S.D. (95% C.I.)	Mean (95% C.I.)	S.D. (95% C.I.)	S.E. (95% C.I.)
2.80 (2.72, 2.87)	0.71 (0.66, 0.77)	-4.83 (-4.94, -4.72)	0.50 (0.42, 0.60)	-5.74 (-5.87, -5.61)	0.91 (0.83, 1.00)	2.75 (2.56, 2.95)

Table 2 Provides estimated mean and standard deviation (S.D.) values together with approximate 95% confidence intervals for the parameter distributions of the model described in the main paper. Estimate of the within group standard error (S.E.) of the residuals with 95% CI is also provided.

Docetaxel – Forwards Alignment

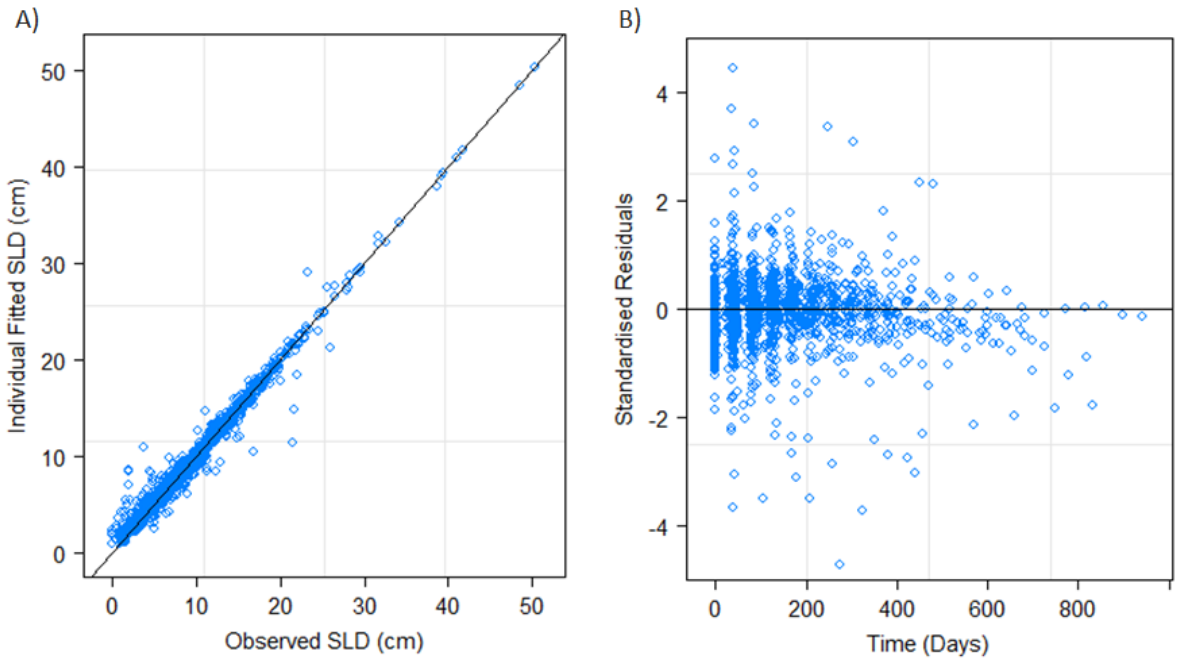

Figure 3 Panel A shows the observed SLD values against the individual fitted. Panel B shows the individual standardised residuals over time.

bioRxiv preprint doi: https://doi.org/10.1101/109934; this version posted February 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

log(A)		log(B)		log(C)		Residual
Mean (95% C.I.)	S.D. (95% C.I.)	Mean (95% C.I.)	S.D. (95% C.I.)	Mean (95% C.I.)	S.D. (95% C.I.)	S.D. (95% C.I.)
2.09 (2.03, 2.15)	0.59 (0.55, 0.64)	-5.27 (-5.38, -5.16)	0.59 (0.50, 0.69)	-5.97 (-6.07, -5.86)	0.80 (0.72, 0.89)	1.06 (1.00, 1.11)

Table 3 Provides estimated mean and standard deviation (S.D.) values together with approximate 95% confidence intervals for the parameter distributions of the model described in the main paper. Estimate of the within group standard error (S.E.) of the residuals with 95% CI is also provided.

Docetaxel – Reverse Alignment

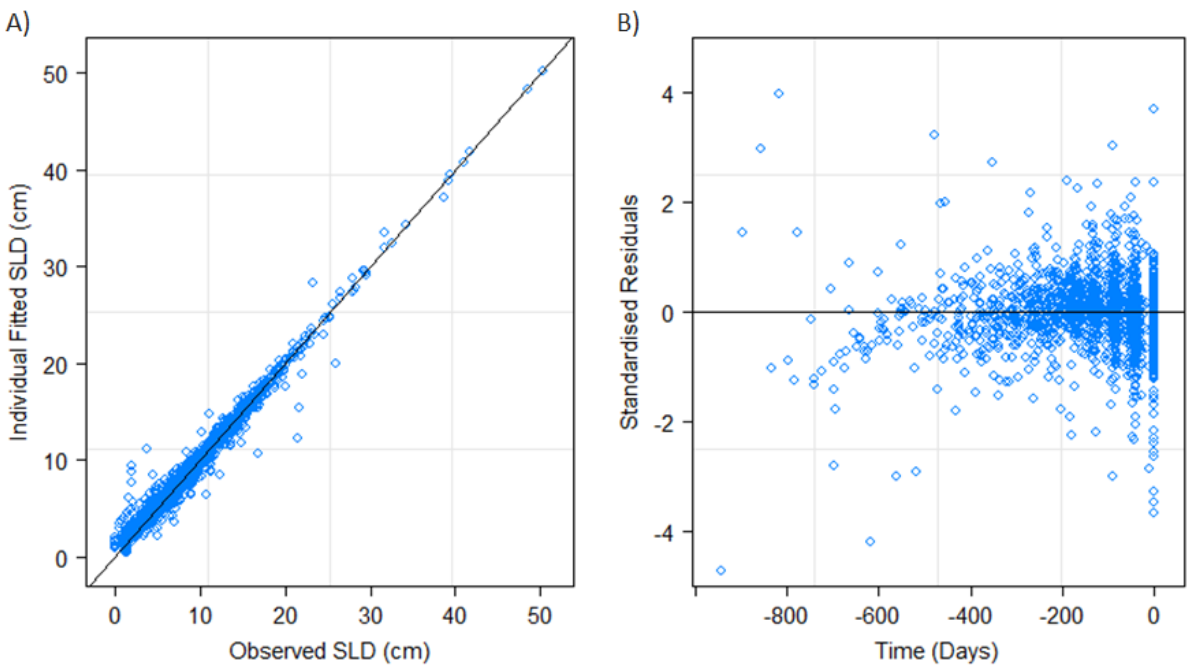

Figure 4 Panel A shows the observed SLD values against the individual fitted. Panel B shows the individual standardised residuals over time.

bioRxiv preprint doi: https://doi.org/10.1101/109934; this version posted February 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

log(A)		log(B)		log(C)		Residual
Mean (95% C.I.)	S.D. (95% C.I.)	Mean (95% C.I.)	S.D. (95% C.I.)	Mean (95% C.I.)	S.D. (95% C.I.)	S.D. (95% C.I.)
2.04 (1.98, 2.11)	0.67 (0.62, 0.72)	-5.17 (-5.26, -5.08)	0.47 (0.40, 0.56)	-5.95 (-6.04, -5.87)	0.55 (0.50, 0.61)	1.06 (1.02, 1.12)

Table 4 Provides estimated mean and standard deviation (S.D.) values together with approximate 95% confidence intervals for the parameter distributions of the model described in the main paper. Estimate of the within group standard error (S.E.) of the residuals with 95% CI is also provided.

Paclitaxel/Carboplatin – Forwards Alignment

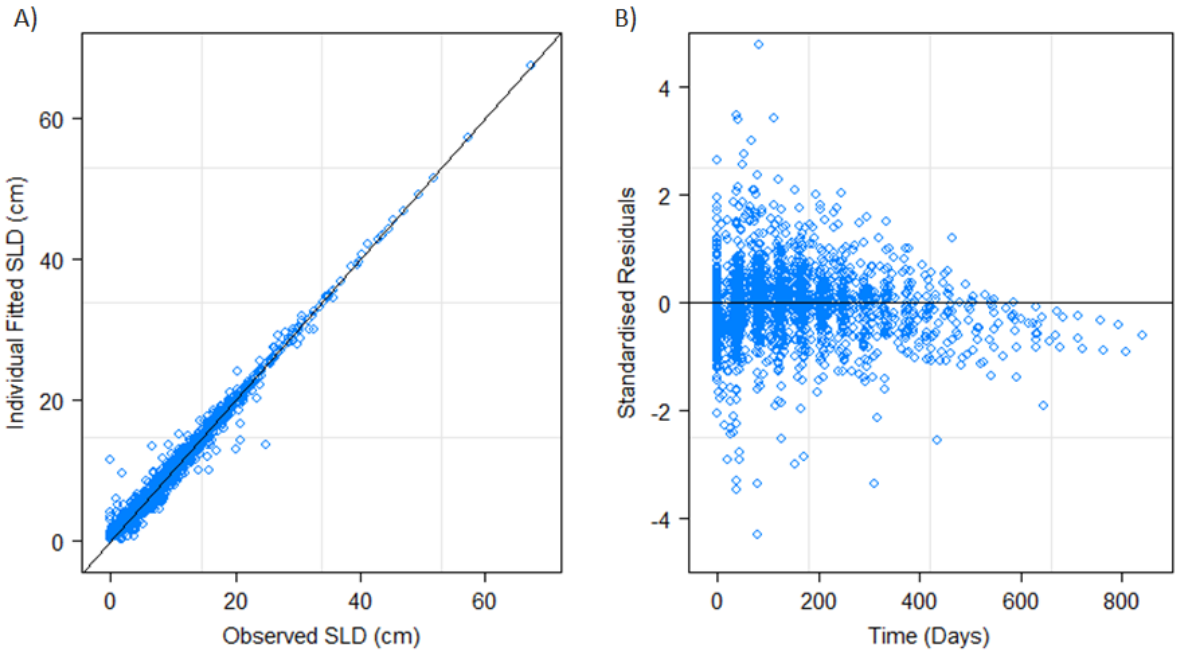

Figure 5 Panel A shows the observed SLD values against the individual fitted. Panel B shows the individual standardised residuals over time.

bioRxiv preprint doi: https://doi.org/10.1101/109934; this version posted February 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

log(A)		log(B)		log(C)		Residual
Mean (95% C.I.)	S.D. (95% C.I.)	Mean (95% C.I.)	S.D. (95% C.I.)	Mean (95% C.I.)	S.D. (95% C.I.)	S.D. (95% C.I.)
2.40 (2.34, 2.47)	0.68 (0.63, 0.73)	-4.51 (-4.61, -4.42)	0.85 (0.78, 0.93)	-6.27 (-6.36, -6.18)	0.76 (0.69, 0.84)	1.21 (1.16, 1.26)

Table 5 Provides estimated mean and standard deviation (S.D.) values together with approximate 95% confidence intervals for the parameter distributions of the model described in the main paper. Estimate of the within group standard error (S.E.) of the residuals with 95% CI is also provided.

Paclitaxel/Carboplatin – Reverse Alignment

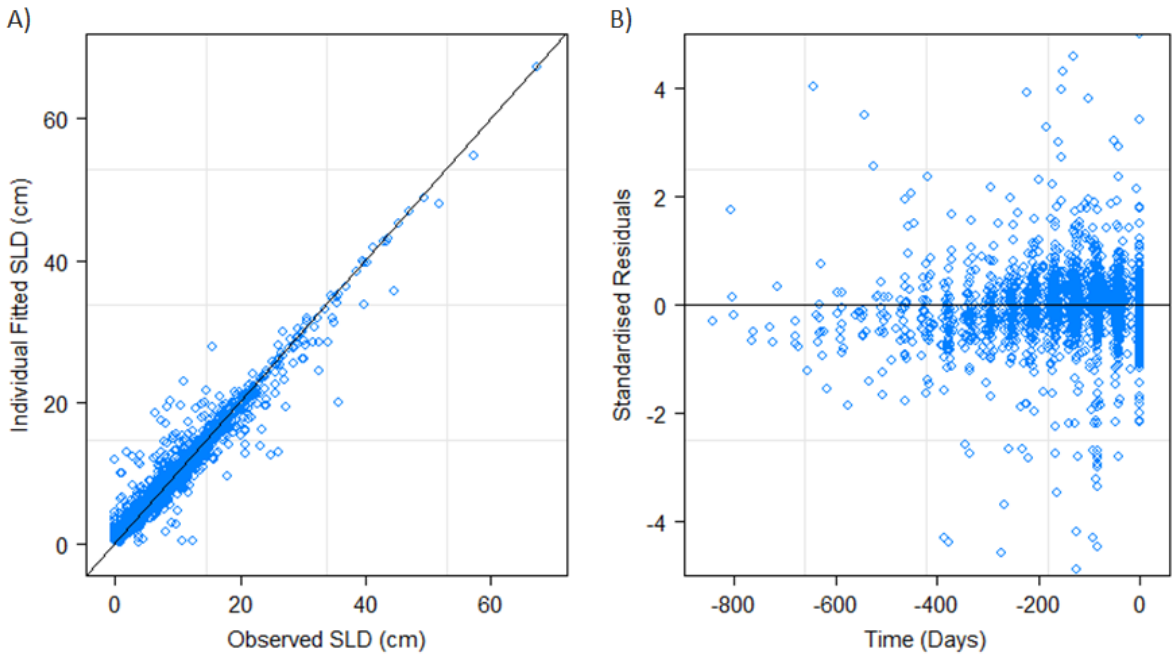

Figure 6 Panel A shows the observed SLD values against the individual fitted. Panel B shows the individual standardised residuals over time.

bioRxiv preprint doi: https://doi.org/10.1101/109934; this version posted February 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

log(A)		log(B)		log(C)		Residual
Mean (95% C.I.)	S.D. (95% C.I.)	Mean (95% C.I.)	S.D. (95% C.I.)	Mean (95% C.I.)	S.D. (95% C.I.)	S.D. (95% C.I.)
1.94 (1.86, 2.02)	0.79 (0.73, 0.86)	-4.65 (-4.76, -4.54)	0.51 (0.40, 0.65)	-5.33 (-5.41, -5.26)	0.67 (0.62, 0.73)	2.07 (1.99, 2.16)

Table 6 Provides estimated mean and standard deviation (S.D.) values together with approximate 95% confidence intervals for the parameter distributions of the model described in the main paper. Estimate of the within group standard error (S.E.) of the residuals with 95% CI is also provided.