Survival analyses in lung cancer

Kari Chansky¹, Dragan Subotic², Nathan R. Foster³, Torsten Blum⁴

¹Cancer Research and Biostatistics, Seattle, USA; ²Clinic for Thoracic Surgery, Clinical Center of Serbia, University of Belgrade School of Medicine, Belgrade, Serbia; ³Division of Biomedical Statistics and Informatics, Mayo Clinic and Mayo Foundation, Rochester, Minnesota, USA; ⁴Lungenklinik Heckeshorn, Belin, Germany

Correspondence to: Prof. Dragan Subotic, MD, PhD. Clinic for Thoracic Surgery, Clinical Center of Serbia, University of Belgrade School of Medicine, 26, Koste Todorovica, 11000 Belgrade, Serbia. Email: profsubotic@gmail.com.

Background: Although survival analyses represent one of the cornerstones in oncology in general, some aspects of the reported survival data in lung cancer patients are still not fully elucidated.

Methods: After having defined several open questions, an evidence based approach was applied in order to answer these questions. Areas of interest were: (I) possible uncertainties in reported survival data; (II) survival surrogates; (III) recommended methods for evaluating progression free survival (PFS) as a surrogate endpoint in future datasets; (IV) postoperative lung cancer recurrence and survival.

Results: In recent years, PFS has seen increasing use as a primary endpoint, particularly in phase II trials. This article focuses on the statistical aspects, and particularly on evaluating the ability of PFS to accurately predict the overall survival (OS) outcome. If the data are available from randomized trials, then the evaluation of trial level surrogacy should be carried out, in addition to the methods described in the paper. If it is not a case, the patient-level methods should be applied. Suggestions for "landmark analysis" are also given: (I) classify your cases according to progression status (progressed, progression-free, or unknown) at one or more time points of interest; (II) perform a separate Cox proportional hazards regression analysis for each time point; (III) determine and report the landmark time point where progression status best predicts survival according to the hazard ratios and P values; (IV) calculate the concordance index for each landmark analysis model. The concordance index (or "c-Index") is essentially the probability that for any two randomly selected cases, the case that is predicted to have the worst outcome, does in fact have the worst outcome.

Conclusions: the widening spectrum of diagnostic and treatment in pulmonary oncology imposes the need for an updated knowledge about statistical method that would fit best for the analysed problem.

Keywords: Lung cancer; survival; progression free survival (PFS); surgery

Submitted Oct 01, 2016. Accepted for publication Nov 03, 2016. doi: 10.21037/jtd.2016.11.28 View this article at: http://dx.doi.org/10.21037/jtd.2016.11.28

Introduction

Like in other types of cancer, in the lung cancer survival analysis, one might be caught into pitfall to get the impression about survival of operated or non-surgically treated patients as a percentage of patients alive at the end of the study period. This simple measure is informative only if all of the patients were observed for the same length of time, which is usually not a case.

There are two main methods of survival analysis: Kaplan-Meier and life-table method. Both methods operate with uncensored and censored cases, the first term relating to patients who are observed until the endpoint of interest (e.g., recurrence or death) and the second to those who survive beyond the end of the follow-up or who are lost to followup at some point.

In the life table method, the total observation period is divided into fixed intervals, usually months or years. After the calculation of proportion of patients surviving till the end of the interval, for each subsequent interval, a cumulative survival is calculated. For example, if the percent of the patients surviving the first interval is 90% and is 80%

3458

for the second and third intervals, the cumulative survival percentage is 57.6% ($0.9 \times 0.8 \times 0.8 = 0.576$).

The Kaplan-Meier methods differ from the previous one in a way that the proportion of patients surviving to each point that a death occurs is calculated, not at fixed intervals. As a consequence, the stepwise changes in the cumulative survival rate appear to occur independently of the intervals on the "X" axis.

Multiple factors express influence to survival and many of them vary on the interval scale, like age, number of analysed nodes and/or node groups, biochemical analyses or markers. Dividing the analysed group according to each interval value would cause the excessive decrease of the number of subjects for each analysis for meaningful conclusions, together with many curves and possible comparisons that would compromise the interpretation. In that situation, the usually used method is Cox proportional hazards regression model, which estimates the influence of multiple variables on survival from data including censored observations. That is not possible with conventional multiple regression analysis, that cannot deal with censored observations.

Possible uncertainties in reported survival data

In the survival analysis, there may be a significant proportion of patients included at the end of the study period, with a follow up, for example of only 1 *vs.* 10 years follow up of earliest patients. So, one of the clinicians' questions could be: in which way these lately included patients affect the obtained results (survival rate) and how high their percentage should be?

The answer to this question of course depends upon the typical survival time of the patients in your population, relative to the length of follow-up for the more recently accrued patients (as well as for the patients that were accrued early but subsequently lost to follow-up). A censored observation is not completely ignored, but only provides partial information toward the survival estimates, and censored observations do not contribute to the "power" of an analysis. A large number of censored observations, which will appear as tick marks near the left end of the survival curve if censoring is shown, will result in instability of the survival estimates. If analyses are re-run at a later date, with further follow-up and events occurring in these patients, the new estimates may be substantially different from what was initially seen. A rule of thumb for clinical trial planning is that your observation period after the last patient is accrued should be at least as long

as the expected median survival for your population [or the median progression free survival (PFS), if PFS is your primary endpoint]. A commonly reported metric is the median follow-up time among patients that were alive at last contact. In study populations with lengthy expected survival, this issue is one argument for using PFS, with its shorter failure times, as a surrogate endpoint.

If the analysis of prognostic factors in 5-year survivors after surgery is planned (1), one of the questions could relate the preferred method-life table or Kaplan-Meier? In other words, after five years, which aspects of survival and prognostic factors analysis are susceptible to the influence of the applied survival analysis method? Should the zero time be the date of surgery, or five years postoperatively?

In the analysis of a subset of patients that are 5-year survivors after surgery, the zero time should be set at five years after surgery, and not the date of surgery. Formal comparisons-P values and hazard ratios, will be affected by the choice of zero time. One fairly obvious issue is that when using Cox proportional hazards regression analysis, and using the surgery date as time zero with survival curves not separating between groups until 5 years, the proportional hazards assumption is clearly violated. The log-rank tests based on the Kaplan-Meier estimations are also affected. Power of the log-rank test is optimized when hazards in the comparator populations are proportional. With the zero-time set up on the day of surgery, the reported P values for the comparisons will be lower than if time zero was chosen appropriately. Clearly, absolute differences in survival times are smaller relative to the overall survival (OS) time of the group as a whole.

Aside from choosing the appropriate zero time, estimates of OS such as Kaplan-Meier are not the best choice in the presence of competing risks (in this case, death due to a cause other than cancer is a competing risk when evaluating the time to death due to cancer). An analysis of cumulative incidence, rather than failure-free survival, would better handle this situation (2).

Survival surrogates

OS is the "gold standard" endpoint for the majority of clinical trial settings in oncology (3). However, in recent years, PFS has seen increasing use as a primary endpoint, particularly in phase II trials, for several reasons (4). Trials utilizing PFS as the primary endpoint can reach completion more quickly as compared to trials with OS as the primary endpoint, because disease progression is typically observed

Journal of Thoracic Disease, Vol 8, No 11 November 2016

at an earlier time point than death. Moreover, with advances in second- and third-line therapy and supportive care, differences in treatment effect between competing treatment regimens can be obscured or even confounded by the choice of treatment post-progression, a factor that is seldom controlled by the clinical trial protocol.

PFS

The performance of PFS and other endpoints as a surrogate for OS varies among different disease and treatment settings. There are multiple considerations regarding the suitability of a surrogate endpoint, however this article focuses on the statistical aspects, and particularly on evaluating the ability of PFS to accurately predict the OS outcome. The topic has received considerable attention in the thoracic oncology literature, with various methods employed. Here we review four different exemplary studies of three different lung cancer disease and treatment settings. These four studies all employed correct techniques. Examples abound of studies using incorrect techniques, such as the computation of a simple correlation coefficient between the PFS outcome and the OS outcome. Such a technique does not account for the censored nature of survival and PFS data.

For those who have access to data, we recommend methods that can be used to evaluate the relationship between PFS and OS in your own database. The suggested methods are most suitably applied to the data from clinical trials, however the analysis of certain detailed institutional series or other databases, though not ideal, is possible. The aim of this paper is to recommend methods that achieve a balance between ease of use and correct technique.

Past evaluations of PFS as a surrogate endpoint

Advanced non-small cell lung cancer (NSCLC)

Mandrekar *et al.* (5) pooled the data from four North Central Cancer Treatment Group (NCCTG) phase II trials, conducted between 2001 and 2007, to examine PFS and tumour response endpoints with respect to their ability to predict OS. PFS and response endpoints were separately modelled as time dependent covariates in Cox regression models on OS, stratified by trial and adjusted by other known baseline prognostic factors. Landmark analyses were then performed evaluating progression (and response) status at 8, 12, 16, 20 and 24 weeks post-registration. The ability of each endpoint to discriminate patients with different survival times was evaluated by calculating a concordance index (6) in conjunction with landmark analyses, for each of the endpoints in question. For all analysis strategies, PFS was the superior surrogate endpoint, in terms of successfully predicting OS outcomes. Patients who had progressed at any time during treatment showed much worse OS, and this effect was more dramatic than the improved prognosis observed in patients with a partial or complete response. In landmark analyses, PFS also outperformed indicators of treatment response, with the highest concordance index being associated with the 12 week landmark.

Extensive stage small cell lung cancer

Foster *et al.* (7) evaluated PFS (along with measures of tumor response) as a surrogate endpoint in first-line extensive SCLC using pooled, patient-level data from nine NCCTG trials that accrued between 1987 and 1999. Analyses at the individual patient level were similar to those described above for NSCLC, with PFS and other potential surrogates modelled as a time dependent covariate, and with landmark analyses at 2, 4, and 6 months. PFS was found to be the best surrogate for OS, with the strongest associations for progression status at 4 and 6 months. In addition, triallevel surrogacy measures were calculated for PFS and response endpoints across the subset of three randomized trials that were included in this study.

In the clinical trial setting, examination of trial level surrogacy is considered to be the best approach for the validation of a surrogate endpoint. These surrogacy measures quantify the association between the treatment effects on OS and the treatment effects on the surrogate endpoints. Trial-level surrogacy was measured in multiple ways, including recommended conventional methods (8). The association between treatment effects on OS and the surrogate endpoints of PFS and response was evaluated by calculating the Spearman's rank correlation coefficient, along with the R² value from a weighted linear regression model (WLS R^2), with weights equal to the sample size of the unit (treating center) from which the data were derived. The treatment effects within each unit were estimated by calculating the log hazard ratios (HRs) and log odds ratios (ORs) from Cox PH and logistic regression models, respectively, depending on the nature of the endpoint. In addition, the surrogacy of the time-to-event putative surrogate endpoint of PFS was quantified by a formal trial-level surrogacy measure, known as the Copula R^2 . Copula R^2 is estimated from a bivariate survival model which models the putative surrogate endpoint and the true endpoint jointly. Both the WLS R^2 and the Copula R^2 value range from 0 to 1, with values close to zero suggesting poor surrogacy, and values close to 1 indicating high surrogacy.

The NCCTG findings in SCLC were later validated in a follow-up study (9) utilizing trial data from a much larger pooled group of ten randomized trials which accrued in the United States cooperative groups and the Japan Cooperative Oncology Group (JCOG), from 1982–2007. This study assessed the surrogacy of PFS with OS at the patient and trial level, with methods similar to the previous analysis. Trial-level surrogacy was assessed through association of the log hazard ratios on OS and PFS across trials, including weighted (by trial size) least squares regression (WLS R^2) of Cox model effects and correlation of the copula effects (copula R^2). The methods regarding these calculations are described by Renfro *et al.* (10). The results of the previous study were validated, with PFS having strong surrogacy for OS at both the patient level and the trial level.

Multimodal NSCLC

The Surrogate Lung Project Collaborative Group (a joint European and North American effort) used techniques similar to those used by Foster *et al.* described above, to evaluate the PFS endpoint in various randomized trials of adjuvant therapy, sequential or concurrent chemotherapy, and modified radiation therapy in locally advanced disease. According to their findings, PFS ranged from good, through very good, to excellent as a surrogate for OS on both the trial level and the independent patient level, depending upon the treatment setting (11).

Recommended methods for evaluating PFS as a surrogate endpoint in future datasets

Clinical trials provide the ideal setting for evaluating surrogate endpoints such as PFS. Clinical trials provide the framework for relative homogeneity in the patient population, uniform methods for response assessments, accurate determinations of progression times, and consistent treatment delivery. The "gold standard" method described above, the evaluation of trial level surrogacy, can only be employed with access to data from multiple randomized clinical trials. If the data are available from a collection of randomized trials, then the evaluation of trial level surrogacy should be carried out, in addition to the methods described below. Although it is more difficult to employ, trial-level surrogacy is the most powerful method for evaluating surrogate endpoints. However, if the data in question are not from randomized trials, which are often the case, then only the patient-level methods described below should be applied.

The methods described by Foster *et al.* (9) that were used to evaluate the large, pooled, international cooperative group dataset are the best approach when the data are available from randomized trials. However in the event that randomized trial data are not available, the methods used in the evaluation of the NCCTG lung cancer data (5) are exemplary in their correct implementation, applicability, and ease of use and interpretation. Our recommendations for single-arm trials or other series data, outlined below, are therefore based on the most easily employed methods described in that paper.

Defining and determining the progression-free survival endpoint

In any case, the PFS endpoint must be clearly defined and consistently measured. For clinical trials, PFS time is defined as the date of trial registration until objective disease progression or death, whichever occurs first (12). Patients alive and progression free at last contact are censored observations. In contrast, time to progression (TTP) is defined simply as the time between trial registration and objective tumor progression. Death is not part of the endpoint; a patient that has died prior to progression counts as a censored observation. Another endpoint that is occasionally used is time to treatment failure (TTF), where failure time is the time at which first-line treatment is terminated for any reason, including progression or adverse events. The relative merits of PFS and TTP are discussed from the regulatory standpoint in government agency publications from the United States and Europe (European Medicines Agency, 2013, U.S. Department of Health and Human Services, Food and Drug Adminstration, 2007). The PFS endpoint could be adapted to the hospital series setting by considering the date of starting planned treatment as the start time

Regardless of whether the investigator is evaluating the data from a series or a clinical trial, attention should be paid to the method and timing of determination of disease progression. Progression times are sensitive to the disease assessment interval, which should therefore be standardized across the patient population. Methods have been devised to deal with the width of assessment intervals (13) relative to progression times, but ideally the assessment interval should not vary much between patients in the cohort. Although this issue is particularly problematic in the comparison

Journal of Thoracic Disease, Vol 8, No 11 November 2016

of two treatment arms with differing cycle length, in the case of evaluating PFS in a single treatment population, widely varying assessment intervals will introduce excess variation in PFS measurements. To further ensure accuracy and consistency, the method of assessment for tumor measurement (imaging technique) should be consistent between baseline and follow up, and an objective response and progression determination method, such as RECIST (14), should be used to determine progression.

Data analysis

As mentioned, for randomized trial data, we recommend a thorough study of the evaluation of trial level surrogacy. Otherwise, in the case of single arm trials or other data, at the very least follow these steps to conduct a "landmark analysis" in conjunction with the calculation of a concordance index:

- (I) Classify your cases according to progression status (progressed, progression-free, or unknown) at one or more time points of interest. (e.g., 2, 4 and 6 months). Progression status for a given time point is only known if the patient has progressed prior to the time point, or the date of last contact for follow-up is after the time point. (If the patient has died without documented progression, then the progression status is "progression."
- (II) Perform a separate Cox proportional hazards regression analysis for each time point, where the explanatory variable is progression status at that time point, and the outcome variable is OS. You may stratify your analyses by any known prognostic factors, such as weight loss or performance status. Include only those cases that could be classified. Exclude these with progression status for the time point. Note: if follow-up is complete and mature, and the time points are chosen appropriately to work with disease assessment times, there should not be a large number of excluded cases.
- (III) Determine and report the landmark time point where progression status best predicts survival according to the hazard ratios and P values from the proportional hazards analyses. Draw Kaplan-Meier survival curves according to progression status (yes *vs.* no) for each time point.
- (IV) Calculate the concordance index for each landmark analysis model. The concordance index (or "c-Index") is essentially the probability that for any two randomly selected cases, the case that is predicted to have the worst outcome (by the model

parameters according to the predictor variable(s) in question which is in this case progression status at "X" months) does in fact have the worst outcome. A C-index of 0.5 would indicate that outcomes are random with respect to the prediction rule (progression status at X months), and a C-index

A C-index of 0.5 would indicate that outcomes are random with respect to the prediction rule (progression status at X months), and a C-index of 1 would indicate a perfect rule. The calculation of a simple C-index involves the evaluation of all possible pairs in the data, excluding those for which it can't be determined which case has the worst outcome in terms of time to death. An example of an excluded pair would be one in which one case is censored (alive at last contact) at a time that is less than the time of death for the other case. It cannot be determined which patient survived (or will survive) the longest. In its simplest form (Harrell, 1982), Harrell's C is easily calculated and available via several different R packages (Appendix 1).

Improvements have been proposed to the basic c-index, in order to avoid the issue of pairs made "unusable" by censoring. These improvements are based on the actual survival models rather than the observations. Gönen and Heller (15) developed a "model-based" analog to the c-index, called the Concordance Probability Estimate (CPE), which effectively estimates the concordance index from the model parameters rather than the observed outcomes, thus avoiding the exclusion of non-informative pairs due to censoring. If this method is desired, R packages are also available and easy to use. An added advantage of this technique is that one can calculate the percent improvement in the CPE of when you add progression status to a "base model", which may only include your known baseline prognostic variables.

Some additional clarifications

Despite sufficient evidence helping clinicians to adequately select the appropriate statistical method, some aspects may still remain unclear.

(I) Is there any difference in rationale for the use of PFS in studies of early stages NSCLC *vs.* advanced stage?

As previously mentioned, populations with longer survival times require longer observation periods for OS. So PFS or disease free survival (DFS) may be the more desirable endpoint in early stage and/or resected NSCLC, especially if the investigator is unable to plan for a lengthy follow-up period. More work is needed to evaluate the suitability of PFS as a surrogate for OS in early stage disease. (II) What are the problems related to the treatment response assessment in studies using a PFS as a surrogate for OS?

Currently, calculation of the PFS is based on unidimensional measurements of the tumour size by using the Response Evaluation Criteria in Solid Tumours (RECIST) system. An important concern that relates to this way of the treatment response assessment is that criteria for disease progression (20% increase in tumour diameters) may not be met in some situations. This because the effect of some new cytostatic or targeted treatment protocols may not lead to tumour shrinkage, but to stable disease or tumour texture change as well. Furthermore, by using RECIST criteria, it is not easy to assess the treatment response at the level of pleura or pericardium.

Volumetric assessment of the tumour change is a valuable alternative, extending the analysis to the tumour density as well. Volumetric measurements were found useful as an early marker of treatment response in NSCLC for detection of subtle changes in indolent disease (16-18). However, unlike for RECIST, a widely accepted definition of response has not yet been established for volumetric change.

Postoperative lung cancer recurrence and survival

The influence of intensified postoperative follow-up on survival or local recurrence detection could not be clearly demonstrated. In around 50-67% of patients, recurrence will appear before a scheduled control because of the onset of symptoms (19). As there are not many studies addressing that point, the evidence-based explanation is not available. In the analysis of this type of survival, lead- and length-time biases should be kept in mind, like in the lung cancer screening. To remind, the lead/time bias causes the impression of prolonged survival in the screened vs. non/ screened group, because of the earlier tumour detection in the screened group, even in the absence of symptoms. Even in case of failure of the initiated treatment, with the same time point of death in both groups, the survival of patients in the screened would be longer owing to the interval between the time of tumour detection by screening and of the symptom onset in the non-screened group. Length time bias is another type of false impression of prolonged survival, caused by more indolent tumours in one group compared to another. In the group with more indolent tumours, more patients with tumour will be detected vs. group with more fast growing tumours. The prolonged survival in the group with more indolent tumours cannot be

attributed to the early initiated treatment.

One of the questions in this field could relate the preferred method for survival analysis in limited series of quite rare procedures, like for example, completion pneumonectomy for postoperative lung cancer recurrence? It is sometimes necessary to run a multicentric study to collect around 30–40 cases. What are potential pitfalls of two survival curves—one from the date of the first, and another from the date of redo surgery?

In this situation, the population is defined by the occurrence of a second surgery due to recurrence. This is similar to something we are seeing in the literature called post-progression survival. In some settings, postprogression survival appears to be more related to OS than PFS, and is therefore a topic that is receiving some attention.

As for pitfalls of using date of original surgery as time zero, similar to example above, there will necessarily be a period of no failures and no censored observations prior to the earliest time when second surgery occurs. So all of the previously described pitfalls exist in this example. The proportional hazards assumption is violated, and P values for comparisons are likely to be artificially smaller.

Pitfalls of using time of redo surgery as time zero: different from the example above, the time between original surgery and redo surgery differs between patients. So if you look at survival from date of redo surgery, the survival outcome of patients with a long period between original and second surgery could be biased downward relative to patients with an earlier re-do, even if survival from original surgery is the same. However, we do frequently see the analogous situation in second line trials in systemic therapies, where recurrence/progression after initial therapy varies with respect to timing, and time zero for the second line trials is at the start of second line treatment. This is the accepted approach, and in fact the findings are often such that the patients with a longer period between completion of first line therapy and recurrence/progression also have the better post-progression prognosis.

Acknowledgements

None.

Footnote

Conflicts of Interest: The authors have no conflicts of interest to declare.

References

- Okada M, Nishio W, Sakamoto T, et al. Long-term survival and prognostic factors of five-year survivors with complete resection of non-small cell lung carcinoma. J Thorac Cardiovasc Surg 2003;126:558-62.
- Gooley TA, Leisenring W, Crowley J, Storer BE. Why Kaplan–Meier fails and cumulative incidence succeeds when estimating failure probabilities in the presence of competing risks. In: Crowley J. editor. Handbook of Statistics in Clinical Oncology Marcel Dekker: New York, NY, USA, 2001:513-3.
- U.S. Department of Health and Human Services, Food and Drug Adminstration. Guidance for Industry: Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics. Rockville, MD: U.S. Department of Health 2007.
- Korn RL, Crowley JJ. Overview: progression-free survival as an endpoint in clinical trials with solid tumors. Clin Cancer Res 2013;19:2607-12.
- Mandrekar SJ, Qi Y, Hillman SL, et al. Endpoints in phase II trials for advanced non-small cell lung cancer. J Thorac Oncol 2010;5:3-9.
- Harrell FE Jr, Lee KL, Califf RM, et al. Regression modelling strategies for improved prognostic prediction. Stat Med 1984;3:143-52.
- Foster NR, Qi Y, Shi Q, et al. Tumor response and progression-free survival as potential surrogate endpoints for overall survival in extensive stage small-cell lung cancer: findings on the basis of North Central Cancer Treatment Group trials. Cancer 2011;117:1262-71.
- Shi Q, Renfro LA, Bot BM, et al. Comparative assessment of trial-level surrogacy measures for candidate time-toevent surrogate endpoints in clinical trials. Computational Statistics & Data Analysis 2011;55:2748-57.
- Foster NR, Renfro LA, Schild SE, et al. Multitrial Evaluation of Progression-Free Survival as a Surrogate End Point for Overall Survival in First-Line Extensive-

Cite this article as: Chansky K, Subotic D, Foster NR, Blum T. Survival analyses in lung cancer. J Thorac Dis 2016;8(11):3457-3463. doi: 10.21037/jtd.2016.11.28

Stage Small-Cell Lung Cancer. J Thorac Oncol 2015;10:1099-106.

- Renfro LA, Shang H, Sargent DJ. Impact of Copula Directional Specification on Multi-Trial Evaluation of Surrogate End Points. J Biopharm Stat 2015;25:857-77.
- Mauguen A, Pignon JP, Burdett S, 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. Lancet Oncol 2013;14:619-26.
- Green S, Benedetti J, Smith A, et al. Clinical Trials in Oncology, Third Edition (Chapman & Hall/CRC Interdisciplinary Statistics). CRC Press, 2012.
- Sridhara R, Mandrekar SJ, Dodd LE. Missing data and measurement variability in assessing progression-free survival endpoint in randomized clinical trials. Clin Cancer Res 2013;19:2613-20.
- 14. Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). Eur J Cancer 2009;45:228-47.
- Gönen M, Heller G. Concordance probability and discriminatory power in proportional hazards regression. Biometrika 1995;92:965-70.
- Mozley PD, Schwartz LH, Bendtsen C, et al. Change in lung tumor volume as a biomarker of treatment response: a critical review of the evidence. Ann Oncol 2010;21:1751-5.
- Zhao B, Schwartz LH, Moskowitz CS, et al. Lung cancer: computerized quantification of tumor response--initial results. Radiology 2006;241:892-8.
- Chang V, Narang J, Schultz L, et al. Computer-aided volumetric analysis as a sensitive tool for the management of incidental meningiomas. Acta Neurochir (Wien) 2012;154:589-97; discussion 597.
- Walsh GL, O'Connor M, Willis KM, et al. Is follow-up of lung cancer patients after resection medically indicated and cost-effective? Ann Thorac Surg 1995;60:1563-70; discussion 1570-2.

Appendix

R package to calculate Harrell's c index:

- ◆ Package "pec" Author: Thomas A. Gerds
- https://cran.r-project.org/web/packages/pec/pec.pdf
- R package for calculating CPE with Standard Error (Gonen) from a Cox Model:
 - * Package "CPE," Authors: Qianxing Mo, Mithat Gonen and Glenn Heller
 - https://cran.r-project.org/web/packages/CPE/CPE.pdf