



HAL
open science

Projecting overall survival in health-economic models: uncertainty and maturity of data

Stéphane Roze, Nicolas Bertrand, Lauriane Eberst, Isabelle Borget

► To cite this version:

Stéphane Roze, Nicolas Bertrand, Lauriane Eberst, Isabelle Borget. Projecting overall survival in health-economic models: uncertainty and maturity of data. *Current Medical Research and Opinion*, 2023, 39 (3), pp.367 - 374. 10.1080/03007995.2023.2167442 . inserm-04529148

HAL Id: inserm-04529148

<https://inserm.hal.science/inserm-04529148>

Submitted on 2 Apr 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License



Projecting overall survival in health-economic models: uncertainty and maturity of data

Stéphane Roze, Nicolas Bertrand, Lauriane Eberst & Isabelle Borget

To cite this article: Stéphane Roze, Nicolas Bertrand, Lauriane Eberst & Isabelle Borget (2023): Projecting overall survival in health-economic models: uncertainty and maturity of data, Current Medical Research and Opinion, DOI: [10.1080/03007995.2023.2167442](https://doi.org/10.1080/03007995.2023.2167442)

To link to this article: <https://doi.org/10.1080/03007995.2023.2167442>



Accepted author version posted online: 10 Jan 2023.



Submit your article to this journal [↗](#)



Article views: 4



View related articles [↗](#)



View Crossmark data [↗](#)

Projecting overall survival in health-economic models: uncertainty and maturity of data

Stéphane Roze¹, Nicolas Bertrand², Lauriane Eberst³ & Isabelle Borget⁴

1. Vyoo Agency, 69100 Villeurbanne, France

2. Medical Oncology department, CHU Lille, F-59000 Lille, France , Université Lille 2, F-59000 Lille, France

3. Institut de Cancerologie de Strasbourg Europe, Medical Oncology Department, Strasbourg, France

4. Department of Biostatistics and Epidemiology, Gustave Roussy ; Oncostat, U1018 Inserm, Paris-Saclay University, "Ligue Contre le Cancer" labeled team, Villejuif, France

Correspondence:

Stéphane Roze

Vyoo Agency, 69100 Villeurbanne, France

Tel.: +33617541425

E-mail: sroze@vyoo-agency.com

Transparency

Declaration of funding

No funding to declare.

Declaration of financial/other relationships

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties. Peer reviewers on this manuscript have no relevant financial or other relationships to disclose.

Author contributions

Isabelle Borget and Stéphane Roze participated in the design, calculation, and manuscript preparation. Lauriane Eberst and Nicolas Bertrand participated in the expert panel and manuscript preparation.

Data availability statement

The data informing this study are available publicly through the references in Table 1.

Abstract

Objective: As lifetime horizons are considered for economic evaluations, the Kaplan-Meier (KM) estimate is used to extrapolate survival in cases of immature overall survival (OS) data. This study estimated the error induced by the choice of distribution when extrapolating different levels of OS maturity.

Methods: Fifteen phase 3 trials reporting KM estimates of OS where at least 70% maturity (i.e. 70% of the population had died during follow-up) were included and compared to artificially created truncated data (30% and 50% maturity). Individual patient-data were reproduced using the Guyot algorithm based on digitized KM curves. Parametric survival distributions were fit for each arm in each study, for each maturity level, using the same time horizon (equal to the maximum follow-up). For each KM curve, the best distribution was chosen based on visual inspection, Akaike/Bayesian information criteria, and external validity. Outcomes were measured as life expectancy in months (LM) and life months gained (LMG).

Results: The Weibull (33%), log-logistic (32%) and log-normal (27%) were most often selected as the best fitting distribution. Compared to LM at full maturity, LM was overestimated in 23% and 40% of cases, at 30% and 50% maturity, respectively. Mean

absolute error was 2.12 months at 30% maturity, and decreased to 0.88 months at 50% maturity. When comparing to mature data, the mean percentage of error in LMG was 126.4% and 62.4% at 30% and 50% maturity, respectively.

Conclusion: The extent of OS maturity increases the risk of error when projecting long-term life expectancy for economic models. Even marginal gains in OS maturity result in more accurate estimations and should be considered when developing models.

Keywords

extrapolation, health economic, survival, modelling, oncology

Short title: Data maturity for projecting survival in economic models

Introduction

Health economic evaluations consist of a comparative analysis of treatment options, both in terms of costs and benefit, expressed as life years gained (LYG) or in quality-adjusted life years (QALY). Extrapolating survival to capture the life course of a disease is recommended by multiple health-technology assessment (HTA) authorities for economic evaluations of oncology drugs, in order to best assess the long-term consequences of the compared strategies [1-3].

Overall Survival (OS) is considered the gold standard as a primary endpoint of randomized controlled trials of interventions aimed at extending life, particularly in oncology [4,5]. A literature review found that this information was not available at the time of approval for more than half of the drugs receiving market authorization by the European Medicines Agency.[6] Oncology drugs in particular benefit from early phase approvals (accelerated procedures, adaptive pathways, etc.) and/or are approved with relatively short trial follow-up, to accelerate the patient's access to innovation [6,7]. Hence, immature data on OS is frequently used for economic evaluations of oncology drugs.

In cases of immature and, as they often are, censored OS data, the Kaplan-Meier (KM) estimator can obtain estimates of the full survival benefit, allowing this data to then be used in an economic model. Multiple methods may be used for extrapolating data, with parametric distributions the most frequently used [8]. As each distribution has a different functional form, resulting in different survival estimates, extrapolating over much larger time horizons than that observed may result in potentially large differences. According to the Decision Support Unit of the National Institute for Health and Care Excellence [8,9], the selection of the best fitting distribution includes the assessment through multiple criteria : 1) visual inspection, assessing how well each parametric model fits the clinical trial data by how closely it visually follows the KM curve; 2) Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC), which provide a statistical test of the relative fit of alternative models to KM curve; 3) clinical validity, obtained through expert clinical judgement on expected long-term survival for the intervention under consideration; and 4) referring to external evidence, such as general population data.

The fit of a curve is depending on the information used to estimate its fit. Thus, the best fitting curve may differ when the maturity of the data is increased or decreased. This in turn impacts the estimation of life expectancy in months (LM) and life months gained (LMG) between compared treatments. To quantify the difference, this study aimed to estimate the error on LM and LMG due to the choice of distribution, for varying levels of OS maturity.

Materials and methods

Data selection

A targeted literature review was undertaken to identify phase 3 randomized controlled trials in oncology for both solid tumors and haemato-oncology, published in peer-reviewed journals between 2013 and 2017; selected studies had to report OS KM survival curves in at least one arm (intervention or comparator), with a minimum of 70% maturity (defined as the rate of deaths (events) observed during the follow-up period) [10]. The KM curves also had to be technically digitizable, ensuring that the number of patients at risk could be extracted for each KM curve.

Digitalization and extrapolation

In order to test the impact of maturity level on LM and LMG, artificially truncated maturity levels were created at 30% and 50% of maturity. These artificially created truncated data were then compared against the full (>70%) maturity identified in the publication. A three-step process was then performed for each OS KM curve. First, the KM curve was digitized using Digitizeit software (www.digitizeit.de version 2.3.3) as recommended by Guyot *et al* [11]. Individual patient data was then reproduced using the algorithm by Guyot and reconstructing the KM data on which the survival curve was based [11]; this step was done using R version 3.5.1. Finally, parametric survival distributions were fit for each OS KM curve using the R survival package [12]. For each selected study, each of the five distributions (Weibull, Gompertz, exponential, log-logistic and log-normal) were fit for each treatment arm (intervention [I] and comparator[C]) and each maturity level (full maturity, truncation at 50% maturity [50%], and truncation at 30% maturity [30%]). The curves were fit on the same time horizon, *i.e.* to the maximum follow-up of both groups in each study (in months), resulting in a total of 450 distributions for the 15 selected studies. Figure 1 shows an example of data at full maturity, 50% maturity and 30% maturity with conceptualized 95% confidence intervals. Digitalization and extrapolation was performed independently by two health-economists, with a quality check for each individual extrapolation.

Distribution selection process

For each KM curve, the best distribution was chosen by an expert panel of two oncologists and two health economists. The expert panel first assessed fit based first on visual inspection only, followed by Akaike/Bayesian information criteria and finally external validity [8,13]. To prevent board participants from being influenced by knowledge of full maturity, KM curves were presented sequentially first by type of cancer (solid tumor, then other) then by maturity (first 30%, then 50% and finally full). The curves were shown without any additional identifying information to avoid bias. Distributions were compared using the chi-square test.

Statistical analysis

For each distribution based on the 90 KM curves (2 treatment arms x 3 maturity levels x 15 studies), the restricted mean survival time (RMST) [14], representing a patient's life expectancy in months (LM), was estimated via the area under the curve (AUC). The number of cases where RMST was over- or under-estimated for 30% and 50% maturity compared to

full maturity, was then calculated. To quantify the error in terms of LM, the difference between RMSTs at full maturity versus 30% and 50% was then calculated as follows for each treatment arm:

$$ErrRMST_{(full-truncated)} = RMST_{full} - RMST_{truncated}$$

Since the RMST differences (ErrRMST) can be negative or positive, meaning that studies that under-estimate LM can compensate studies that over-estimate LM, we also calculated the absolute error using (AbsErrRMST) the absolute values of RMST difference between full versus 30% and 50% maturity for each treatment arm:

$$AbsErrRMST_{(full-truncated)} = Abs(RMST_{(full-truncated)})$$

As life expectancy magnitude is specific to each study and treatment arm, the relative errors in percentage at 30 and 50% maturity compared to full were calculated for both treatment arms:

$$\%ErrRMST_{(full-truncated)} = \frac{AbsErrRMST_{(full-truncated)}}{RMST_{full}}$$

LMG were calculated for the intervention versus comparator as follows for each study, where x is the maturity:

$$LMG_x = RMST_{intervention,x} - RMST_{comparator,x}$$

The percent error in LMG was estimated as follows:

$$\%ErrLMG_{full-truncated} = Abs\left(\frac{LMG_{full} - LMG_{truncated}}{LMG_{full}}\right)$$

This parameter takes into account various levels of LMG among the 15 studies.

Results

Selected studies

A total of 15 trials that met the inclusion criteria and are summarized in Table 1. Identified studies included various localizations of cancer as well as treatments. Average maximum duration of follow-up was 51 months (median=37 months, range=10-95) and the mean maturity level was 88%.

Distributions

Type of distribution choice

Based on the selection process, overall, the three best fitting distributions were Weibull (33%), log-logistic (32%) and log-normal (27%); the exponential distribution was never chosen. There were no differences observed in the choice of distributions for treatment arms (i.e. the same distribution was chosen for both treatment arms). Further, as shown in Figure 2, there was no difference in the distribution chosen across the three maturity levels ($p=0.65$).

Life expectancy in months

Across both treatment arms, the RMSTs were overestimated in 23% (7/30) and 40% (12/30) of the time, at 30% and 50% maturity, respectively (Figure 3). The mean life month error was 0.58 months for 50% maturity versus full and 1.23 months for 30% maturity versus full. The mean absolute error at 50% maturity was 0.88 months and increased to 2.12 months for 30% maturity (Figure 4). The improvement in the estimation of life expectancy was 2.4 times better when maturity increased from 30% to 50%. The mean (median) percentage error at 30% maturity of RMST was 13.1% (10.4%) and at 50% maturity of RMST was 6.7% (3.3%).

Life months gained

On average, at 30% maturity versus full, the mean percentage of error in LMG was 126.4%; at 50% maturity versus full, the mean percentage of error in LMG was 62.4%. The median percentages were 64.4% and 28.0% at 30% and 50% maturity, respectively. An important heterogeneity was also observed: for 30% maturity, the minimum and maximum were between 8% and 790% and for 50% maturity, it ranged between 0.5% and 377% (Figure 5). Of note, zero percent indicates an absence of error in the estimation of LMG at full maturity for the extrapolations based on 30% and 50% maturity.

Discussion

The objective of this study was to quantify the error in estimated LM and LMG using extrapolations, based on immature OS data often used for pharmacoeconomic evaluations

for HTAs. Compared to OS at full maturity, an underestimation of life expectancy was observed in 77% and 60% of cases, respectively, for 30% and 50% maturity. Further, across the 15 studies, the mean absolute error on life expectancy was reduced by 2.4 times (from 2.12 to 0.88 months) when maturity improved from 30% to 50%. Similar results were observed for LMG where the mean percentage in error was reduced with increasing maturity.

The tendency to underestimate OS with immature data can be explained by the observation of a higher number of deaths during the early the early follow-up of a clinical trial. As such, it follows that extrapolating curves based on premature data (30% maturity) may underestimate OS for rapidly progressive and non-responding patients. Furthermore, extrapolating on immature data does not take into account possible plateaus for long responders. This underestimation of life expectancy is, however, considered conservative in the context of a health economic evaluation. Bullement *et al.* [15] conducted similar work though their work was limited to the scope of immunotherapies in the context of evaluations by NICE. They found a slight under-estimation of projected OS based on available data at time of submission compared to the most recently available data published following submission.

There is also a strong correlation between maturity levels and the associated risk of error. While the maximum duration of follow-up among the 15 selected studies was heterogeneous (between 10 and 90 months), it appeared that the risk of error was higher for slowly progressing diseases, or diseases with an observed plateau where the models did not fit as well. Thus, extrapolations in the context of slower progressing diseases need to be carefully interpreted. Further, in the case of a cancer with either a good likelihood of survival or with slower progression of disease, death can also be due to causes other than cancer, which also explains why survival outcomes can be difficult to predict. Therefore, the natural evolution of the disease along with the maturity of the survival data should be taken into account for accuracy and robustness when considering the extrapolation needs.

While the use of the most mature data is preferable in all cases, it should be noted that there are scenarios where the OS will remain immature when evaluated by the HTA agencies. The first is when OS is immature due to early results because a new drug has a major therapeutic impact or in a disease where there is a high unmet medical need (i.e.

CAR-T cell therapy for the management of large cell B lymphoma [16,17]). To address this, the Commission d'Evaluation Economique de Santé Publique (CEESP) in France has requested updated results when more mature OS data becomes available [18,19]. A second scenario corresponds to pathologies for which multiple lines of treatment are available and for which a benefit in OS is difficult to demonstrate in the short term [20,21]. This is particularly the case for adjuvant therapies in oncology, such as in melanoma, as mentioned by NICE [22]. Finally, a third scenario can arise in diseases with slow progression and/or there is a competing risk of death from another cause. For instance, patients with surgically resectable oesophageal adenocarcinoma can be cured and die from consequences due to long-term consequences or treatment or comorbidities [23]. In both the second and third scenarios, OS may not be the most relevant endpoint. If validated by strong statistical methods [24-26], surrogate endpoints for OS may be a better alternative. This technique was used in the adjuvant treatment of melanoma where OS was estimated using RFS based on a predictive model for nivolumab versus ipilimumab [27]. A correlation ratio between OS and progression-free survival gain has previously been used as a proxy when more mature data were available in one population and then applied to another in a given indication [28]; however, this approach was challenged by NICE [29] given that the validation was based on the use of mature data in a subgroup of people with the BRCA mutation in a trial of olaparib. Previous studies have explored the impact of extrapolations [30-32]. However, to our knowledge, this is the first time that OS has been extrapolated across a large (n=15) sample of select oncology studies avoiding potential technical issues (i.e. digitization process), generating 450 different extrapolations covering a wide range of situations. Further, all selected studies had to be randomized controlled trials with mature OS data, resulting in publications in peer-reviewed, high impact factor journals. As our sample of studies included different types of cancer (both common and rare), at various stages (localized/metastatic) with various interventions (surgery, chemotherapy, immunotherapy, targeted therapy), they offer a good overview of the diversity of clinical situations that can be encountered in the field of oncology. Our process of choosing the best curved based on the combination of statistical and clinical inputs reflects commonly used methods of extrapolating OS in economic evaluations of oncology products for HTA reviews. Applying the same standard methodology to all 15 studies enabled reproducible results while reducing the

heterogeneity of estimations, as opposed to a situation of a meta-analysis based on 15 single and independent experiments. Further, this process highlighted the importance of external clinical validity, and not only statistical criteria of fit, in the selection of survival curves.

Our study has limitations. Our approach assumed that the KM estimate over the first months of mature data is identical to if the KM estimate had been made only with data to this same cut-off. This approach does not reflect the reality that KM estimates change as the data matures. This work was performed following initial evaluations for market authorization or submissions to HTA authorities potentially impacting the clinical experts experience in the given indication and/or the related drug's efficacy. While this may have impacted their choice across possible distributions, the error would have been in the direction of the true OS KM curves. Thus, the presented results may be underestimated. Unlike HTA evaluations where the expert panel is selected based on their specific expertise for a given indication, the two oncologists in this study were selected based for their broad knowledge across all indications. Further, our selection of studies was based on an arbitrary cut-off maturity level of 70%, as a trade-off between the longest-term possible data and the available number of the most recent published studies was needed in order to identify appropriate trials. Use of immature data is standard practice in seeking reimbursement; extrapolation of survival outcomes will nearly always be required since trials rarely report complete survival data. While updating models is likely to reduce uncertainty in decision making, it may not always be possible and there will likely remain uncertainty in long-term outcomes. Future work, including the use of artificial intelligence, should explore alternative methods in reducing the uncertainty around this necessary extrapolation. While our study noted the proportion of models that under- or overestimated overall survival, this was not quantified. Quantifying and exploring the clinically relevant impact of under- or overestimates on overall survival should be explored in future research. Finally, this study does not consider all approaches to extrapolating survival. Flexible extrapolation approaches, including splines or fractional polynomials, are also available and may be appropriate in some circumstances.[9]

This study highlights the impact of OS maturity in reducing error for the extrapolation of KM curves. Extent of OS maturity increases the risk of error when extrapolating for the purpose

of health economic models, with more mature OS data resulting in more accurate estimations. While more mature data in early stages of drug development is often not available, as more mature data becomes available, health economic analyses should be updated and revised. Investigators and HTA agencies should work together to update long-term OS data in order to best inform health economic evaluations of oncology treatments.

Accepted Manuscript

References

1. CADTH. Guidelines for the Economic Evaluation of Health Technologies: 4th Edition. 2017.
2. Haute Autorité de Santé. Choices in methods for economic evaluation - HAS. Methodological Guidance2020.
3. National Institute for Health and Care Excellence. Developing NICE guidelines: the manual. Process and methods [PMG20]. 2014.
4. Driscoll JJ, Rixe O. Overall Survival: Still the Gold Standard: Why Overall Survival Remains the Definitive End Point in Cancer Clinical Trials. *The Cancer Journal*. 2009;15(5).
5. Wilson MK, Karakasis K, Oza AM. Outcomes and endpoints in trials of cancer treatment: the past, present, and future. *Lancet Oncol*. 2015 Jan;16(1):e32-42.
6. Michaeli DT, Michaeli T. Overall Survival, Progression-Free Survival, and Tumor Response Benefit Supporting Initial US Food and Drug Administration Approval and Indication Extension of New Cancer Drugs, 2003-2021. *Journal of Clinical Oncology*. 2022;JCO.22.00535.
7. Booth CM, Del Paggio JC. Approvals in 2016: questioning the clinical benefit of anticancer therapies. *Nat Rev Clin Oncol*. 2017 Feb 20;14(3):135-136.
8. Latimer NR. NICE Decision Support Unit Technical Support Documents. Survival Analysis For Economic Evaluations Alongside Clinical Trials - Extrapolation with Patient-Level Data. London: National Institute for Health and Care Excellence (NICE)
Copyright © 2013 National Institute for Health and Clinical Excellence, unless otherwise stated. All rights reserved.; 2013.
9. Rutherford M, Lambert P, Sweeting M, et al. NICE DSU Technical Support Document 21: Flexible Methods for Survival Analysis. In: Unit NDS, editor. 2020.
10. GebSKI V, Garès V, Gibbs E, et al. Data maturity and follow-up in time-to-event analyses. *International Journal of Epidemiology*. 2018;47(3):850-859.
11. Guyot P, Ades AE, Ouwens MJNM, et al. Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival curves. *BMC Medical Research Methodology*. 2012 2012/02/01;12(1):9.
12. Diaby V, Adunlin G, Montero AJ. Survival modeling for the estimation of transition probabilities in model-based economic evaluations in the absence of individual patient data: a tutorial. *Pharmacoeconomics*. 2014 Feb;32(2):101-8.

13. Woods BS, Sideris E, Palmer SJ, et al. NICE DSU Technical Support Document 19. Partitioned Survival Analysis for Decision Modelling in Health Care: A Critical Review. In: Unit NDS, editor. 2017.
14. Royston P, Parmar MK. Restricted mean survival time: an alternative to the hazard ratio for the design and analysis of randomized trials with a time-to-event outcome. *BMC Med Res Methodol*. 2013 Dec 7;13:152.
15. Bullement A, Meng Y, Cooper M, et al. A review and validation of overall survival extrapolation in health technology assessments of cancer immunotherapy by the National Institute for Health and Care Excellence: how did the initial best estimate compare to trial data subsequently made available? *J Med Econ*. 2019 Mar;22(3):205-214.
16. Neelapu SS, Locke FL, Bartlett NL, et al. Axicabtagene Ciloleucel CAR T-Cell Therapy in Refractory Large B-Cell Lymphoma. *N Engl J Med*. 2017 Dec 28;377(26):2531-2544.
17. Schuster SJ, Bishop MR, Tam CS, et al. Tisagenlecleucel in Adult Relapsed or Refractory Diffuse Large B-Cell Lymphoma. *New England Journal of Medicine*. 2018 2019/01/03;380(1):45-56.
18. Haute Autorité de Santé. Avis d'efficience : KYMRIA[®](Tisagenlecleucel)Lymphome diffus à grandes cellules B 2019. Available from: https://www.has-sante.fr/upload/docs/application/pdf/2019-03/kymriah_ldgcb_15012019_avis_efficience.pdf
19. Haute Autorité de Santé. Avis d'efficience : YESCARTA[®](axicabtagene ciloleucel)Lymphomes diffus à grandes cellules B (LDGCB) et lymphome médiastinal primitif à grandes cellules B (LMPGCB) réfractaire ou en rechute, après au moins deux lignes de traitement systémique 2019. Available from: https://www.has-sante.fr/plugins/ModuleXitiKLEE/types/FileDocument/doXiti.jsp?id=c_2961897
20. Korn EL, Freidlin B, Abrams JS. Overall survival as the outcome for randomized clinical trials with effective subsequent therapies. *J Clin Oncol*. 2011 Jun 10;29(17):2439-42.
21. Saad ED, Katz A, Buyse M. Overall survival and post-progression survival in advanced breast cancer: a review of recent randomized clinical trials. *J Clin Oncol*. 2010 Apr 10;28(11):1958-62.
22. NICE. Appraisal consultation document. Nivolumab for adjuvant treatment of completely resected melanoma with lymph node involvement or metastatic disease. 2018.
23. Alderson D, Cunningham D, Nankivell M, et al. Neoadjuvant cisplatin and fluorouracil versus epirubicin, cisplatin, and capecitabine followed by resection in patients with oesophageal adenocarcinoma (UK MRC OE05): an open-label, randomised phase 3 trial. *The Lancet Oncology*. 2017;18(9):1249-1260.

24. Burzykowski T, Molenberghs G, Buyse M, et al. Validation of Surrogate End Points in Multiple Randomized Clinical Trials with Failure Time End Points. *Journal of the Royal Statistical Society Series C (Applied Statistics)*. 2001;50(4):405-422.
25. Burzykowski T, Molenberghs G, Buyse M. The validation of surrogate end points by using data from randomized clinical trials: a case-study in advanced colorectal cancer [<https://doi.org/10.1111/j.1467-985X.2004.00293.x>]. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*. 2004 2004/02/01;167(1):103-124.
26. Buyse M, Sargent DJ, Grothey A, et al. Biomarkers and surrogate end points--the challenge of statistical validation. *Nat Rev Clin Oncol*. 2010 Jun;7(6):309-17.
27. Suci S, Eggermont AMM, Lorigan P, et al. Relapse-Free Survival as a Surrogate for Overall Survival in the Evaluation of Stage II-III Melanoma Adjuvant Therapy. *J Natl Cancer Inst*. 2018 Jan 1;110(1).
28. Guy H, Walder L, Fisher M. Cost-Effectiveness of Niraparib Versus Routine Surveillance, Olaparib and Rucaparib for the Maintenance Treatment of Patients with Ovarian Cancer in the United States. *Pharmacoeconomics*. 2019 Mar;37(3):391-405.
29. NICE. Final appraisal determination: Niraparib formaintenance treatment ofrelapsed,platinum-sensitive ovarian, fallopian tube and peritoneal cancer 2018. Available from: <https://www.nice.org.uk/guidance/ta528/documents/final-appraisal-determination-document>
30. Hettle R, Posnett J, Borrill J. Challenges in economic modeling of anticancer therapies: an example of modeling the survival benefit of olaparib maintenance therapy for patients with BRCA-mutated platinum-sensitive relapsed ovarian cancer. *J Med Econ*. 2015;18(7):516-24.
31. Bagust A, Beale SJ. Exploring the Effects of Early Censoring and Analysis of Clinical Trial Survival Data on Effectiveness and Cost-effectiveness Estimation through a Case Study in Advanced Breast Cancer. *Med Decis Making*. 2018 Oct;38(7):789-796.
32. Graham CN, Christodoulou A, Knox HN, et al. A within-trial cost-effectiveness analysis of panitumumab compared with bevacizumab in the first-line treatment of patients with wild-type RAS metastatic colorectal cancer in the US. *J Med Econ*. 2018 Nov;21(11):1075-1083.
33. Ascierto PA, McArthur GA, Dréno B, et al. Cobimetinib combined with vemurafenib in advanced BRAF(V600)-mutant melanoma (coBRIM): updated efficacy results from a randomised, double-blind, phase 3 trial. *Lancet Oncol*. 2016 Sep;17(9):1248-60.
34. Carbone DP, Reck M, Paz-Ares L, et al. First-Line Nivolumab in Stage IV or Recurrent Non-Small-Cell Lung Cancer. *New England Journal of Medicine*. 2017 2017/06/22;376(25):2415-2426.

35. Dummer R, Schadendorf D, Ascierto PA, et al. Binimetinib versus dacarbazine in patients with advanced NRAS-mutant melanoma (NEMO): a multicentre, open-label, randomised, phase 3 trial. *Lancet Oncol.* 2017 Apr;18(4):435-445.
36. Fujitani K, Yang HK, Mizusawa J, et al. Gastrectomy plus chemotherapy versus chemotherapy alone for advanced gastric cancer with a single non-curable factor (REGATTA): a phase 3, randomised controlled trial. *Lancet Oncol.* 2016 Mar;17(3):309-318.
37. Gilbert MR, Dignam JJ, Armstrong TS, et al. A Randomized Trial of Bevacizumab for Newly Diagnosed Glioblastoma. *New England Journal of Medicine.* 2014 2014/02/20;370(8):699-708.
38. Grothey A, Van Cutsem E, Sobrero A, et al. Regorafenib monotherapy for previously treated metastatic colorectal cancer (CORRECT): an international, multicentre, randomised, placebo-controlled, phase 3 trial. *Lancet.* 2013 Jan 26;381(9863):303-12.
39. Harbeck N, Huang CS, Hurvitz S, et al. Afatinib plus vinorelbine versus trastuzumab plus vinorelbine in patients with HER2-overexpressing metastatic breast cancer who had progressed on one previous trastuzumab treatment (LUX-Breast 1): an open-label, randomised, phase 3 trial. *Lancet Oncol.* 2016 Mar;17(3):357-366.
40. Kantarjian HM, DeAngelo DJ, Stelljes M, et al. Inotuzumab Ozogamicin versus Standard Therapy for Acute Lymphoblastic Leukemia. *New England Journal of Medicine.* 2016 2016/08/25;375(8):740-753.
41. Motzer RJ, Escudier B, McDermott DF, et al. Nivolumab versus Everolimus in Advanced Renal-Cell Carcinoma. *N Engl J Med.* 2015 Nov 5;373(19):1803-13.
42. Parker C, Nilsson S, Heinrich D, et al. Alpha emitter radium-223 and survival in metastatic prostate cancer. *N Engl J Med.* 2013 Jul 18;369(3):213-23.
43. Seddon B, Strauss SJ, Whelan J, et al. Gemcitabine and docetaxel versus doxorubicin as first-line treatment in previously untreated advanced unresectable or metastatic soft-tissue sarcomas (GeDDiS): a randomised controlled phase 3 trial. *The Lancet Oncology.* 2017;18(10):1397-1410.
44. Stintzing S, Modest DP, Rossius L, et al. FOLFIRI plus cetuximab versus FOLFIRI plus bevacizumab for metastatic colorectal cancer (FIRE-3): a post-hoc analysis of tumour dynamics in the final RAS wild-type subgroup of this randomised open-label phase 3 trial. *Lancet Oncol.* 2016 Oct;17(10):1426-1434.
45. Tewari KS, Sill MW, Long HJ, et al. Improved Survival with Bevacizumab in Advanced Cervical Cancer. *New England Journal of Medicine.* 2014 2014/02/20;370(8):734-743.

46. Zalcman G, Mazieres J, Margery J, et al. Bevacizumab for newly diagnosed pleural mesothelioma in the Mesothelioma Avastin Cisplatin Pemetrexed Study (MAPS): a randomised, controlled, open-label, phase 3 trial. *Lancet*. 2016 Apr 2;387(10026):1405-1414.

Table 1: Characteristics of the selected studies

Primary Author	Year	Population and indication	Intervention	Comparator	Maximum duration of follow-up (in months)	Maturity
Alderson [23]	2017	Surgically resectable oesophageal adenocarcinoma	Cisplatin and fluorouracil	Epirubicin, cisplatin, and capecitabine	95	71%
Ascierto [33]	2016	Previously untreated BRAFV600 mutated unresectable stage IIIC or stage IV melanoma	Cobimetinib + vemurafenib	Placebo +vemurafenib	30	78%

Carbone [34]	201 7	First-line stage IV or recurrent PD-L1 positive	Nivolumab	Platinum- based chemotherap y	28	88%
Dummer [35]	201 7	Advanced unresectable stage IIIC or stage IV NRAS- mutant melanoma	Binimetinib	Dacarbazine	28	91%
Fujitani [36]	201 6	Advanced gastric cancer with a single non-curable factor	Gastrectomy followed by chemotherap y	Chemotherap y alone	79	100%
Gilbert [37]	201 4	Newly diagnosed glioblastoma	Bevacizumab	Placebo	30	82%
Grothey [38]	201 3	Previously treated metastatic colorectal cancer	Regorafenib	Placebo	14	81%
Harbeck [39]	201 6	HER2- overexpressing metastatic breast cancer who had progressed on one previous treatment	Afatinib plus vinorelbine	Trastuzumab plus vinorelbine	43	87%

Kantarjian [40]	2016	Relapsed or refractory acute lymphoblastic leukemia	Inotuzumab ozogamicin	Standard therapy	40	91%
Motzer [41]	2015	Previously-treated advanced clear-cell renal carcinoma	Nivolumab	Everolimus	32	69%
Parker [42]	2013	Castration-resistant metastatic prostate cancer	Radium-223	Placebo	37	94%
Seddon [43]	2017	First-line advanced or metastatic soft-tissue sarcoma	Gemcitabine and docetaxel	Doxorubicin	10	73%
Stintzing [44]	2016	First-line treatment of patients with KRAS exon 2 wild-type metastatic colorectal cancer	FOLFIRI plus cetuximab	FOLFIRI plus bevacizumab	71	99%
Tewari [45]	2014	Recurrent, persistent or metastatic cervical cancer	Chemotherapy with bevacizumab	Chemotherapy without bevacizumab	37	100%
Zalcman [46]	2016	First-line advanced	Pemetrexed plus cisplatin	pemetrexed plus cisplatin	80	94%

		malignant pleural mesothelioma	and bevacizumab			
--	--	--------------------------------------	--------------------	--	--	--

Figure captions

Figure 1: KM curves fit with parametric distributions at full maturity (A), 50% maturity (B) and 30% maturity (C) for both treatment arms

Figure 2: Distribution types for each maturity level

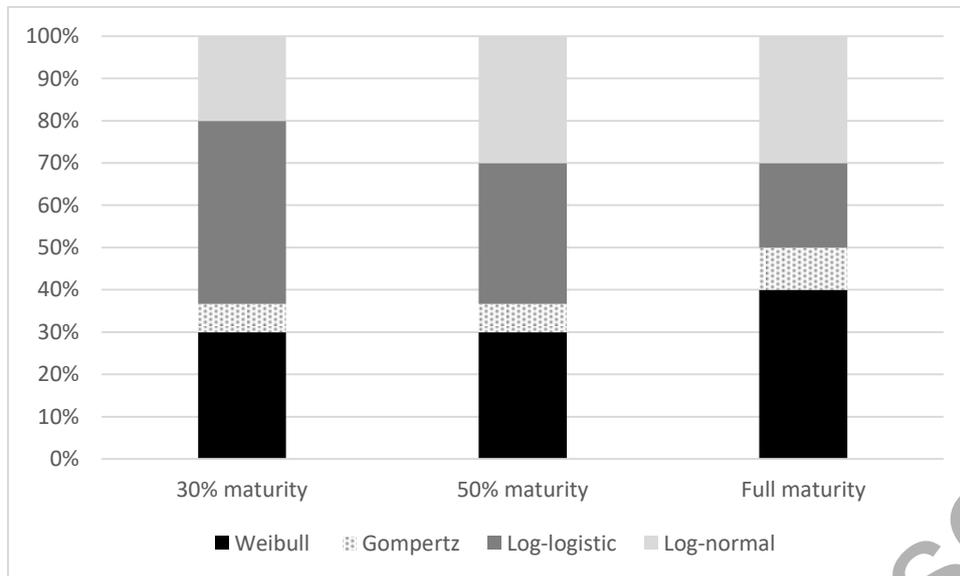
Figure 3: Proportion of cases with over- and underestimation of LM at 30 and 50% maturity

Figure 4: Mean and median of the absolute error between RMST at 30% and 50% maturity versus full (in months)

Figure 5: Percent error in life months gained between 30% and 50% maturity versus full, per study

Accepted Manuscript

Figure 2: Distribution types for each maturity level



Accepted Manuscript

Figure 3: Proportion of cases with over- and underestimation of LM at 30 and 50% maturity

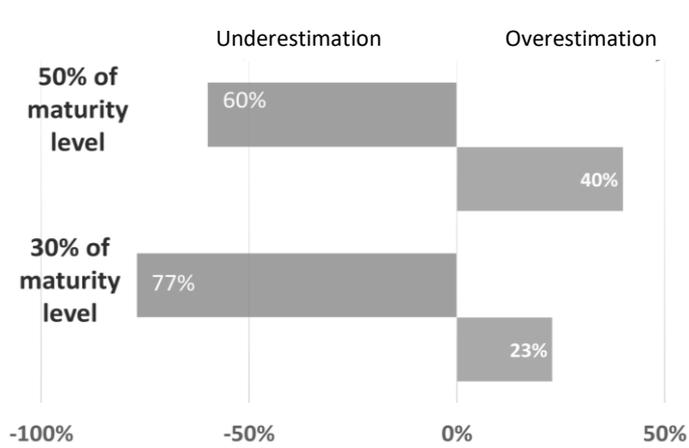


Figure 4: Mean and median of the absolute error between RMST at 30% and 50% maturity versus full (in months)

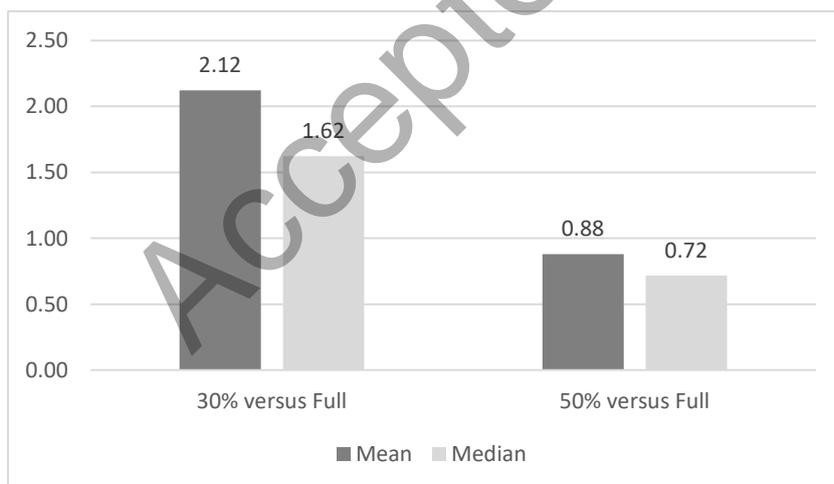
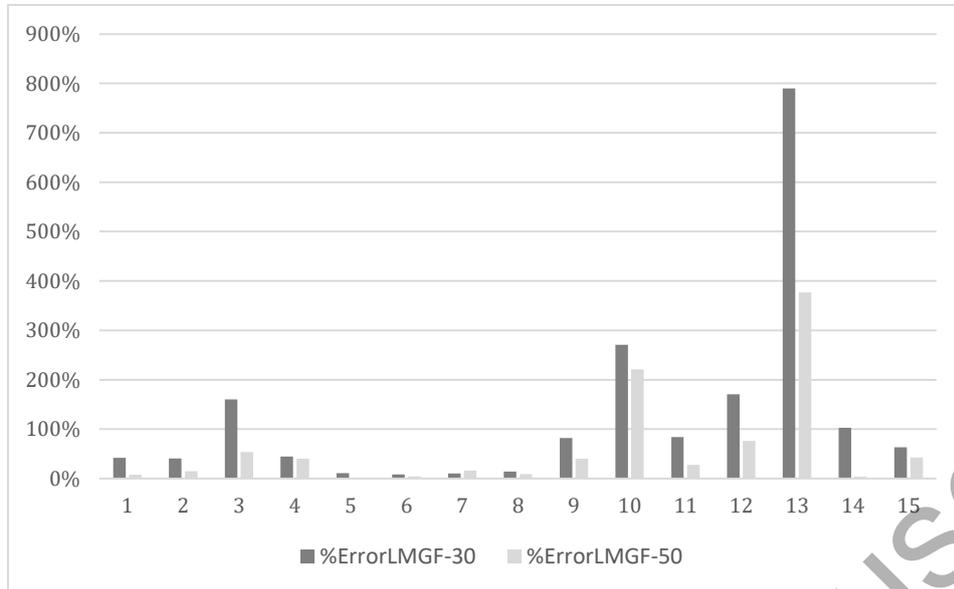


Figure 5: Percent error in life months gained between 30% and 50% maturity versus full, per study



Accepted Manuscript