Objectives: To compare alternative methods for unanchored indirect comparisons of two interventions for overall survival when individual patient data (IPD) is available for both studies (IPD-IPD analyses), and when IPD is only available for one study and aggregate data (AD) for the other study (IPD-AD analyses). ***Methods:*** A case study comparing nivolumab with standard of care (SoC) in 3L small cell lung cancer was performed using the CheckMate032 trial and the Flatiron Health database to compare unanchored comparisons without population adjustment, IPD-IPD adjustment analyses using inverse probability treatment weighting (IPTW), regression adjustment (RA), and doubly robust methods ($DRM_{IPD-IPD}$), and IPD-AD analyses using matching-adjusted indirect comparison (MAIC), simulated treatment comparison (STC), and a novel DRM_{IPD-AD}. Relative treatment effects were expressed with Cox hazard ratios (HRs) and differences in restricted mean survival time (DRMST). A simulation study was performed evaluating the performance of the alternative methods in different scenarios. ***Results:*** Conditional HRs from RA differed from the marginal HRs from IPTW due to non-collapsibility. Therefore, estimates of DRMST in months were used as the relative treatment effect measure for the case study, which provides a collapsible estimand for DR estimates and comparison between methods DRMST with nivolumab versus SoC was higher for RA (3.22) than IPTW (2.72) and $DRM_{IPD-IPD}$ (2.71). MAIC, STC, and DRM_{IPD-AD} estimates were similar (2.88; 2.92; 2.88) but with more uncertainty. The simulations found that when covariates differed substantially between studies, IPTW performed poorly in comparison to RA and DRM. ***Conclusions:*** Analysts should be aware of differences between marginal and conditional treatment effects when performing indirect comparisons of time-to-event outcomes; DRMST may be considered. Regression-based or DRM models may be preferable over other adjustment methods in cases with limited overlap in covariates between studies indirectly compared. Having IPD for both studies is preferable over having IPD for only one.

P44
NOVEL AND EXISTING FLEXIBLE SURVIVAL METHODS FOR NETWORK META-ANALYSES

Heeg B,[1] Garcia A,[2] van Beekhuizen S,[3] Verhoek A,[1] Roychoudhury S,[4] Cappelleri J,[5] Postma M,[6] Ouwens M[7]
[1]*Ingress-health, Rotterdam, ZH, Netherlands,* [2]*Ingress-Health, rotterdam, Netherlands,* [3]*Ingress-Health, Rotterdam, Netherlands,* [4]*Pfizer, Inc., Gladstone, NY, USA,* [5]*Pfizer Inc, Newington, CT, USA,* [6]*University of Groningen, Groningen, Netherlands,* [7]*AstraZeneca, Gothenburg, Sweden*

Objectives: The NICE Decision Support Unit has recently published technical support document 21, focusing on flexible survival methods and considering excess mortality. In this study, we research these flexible survival methods by assessing the effect of treatment on the different parameters of the distributions and the implementation of these flexible excess survival models in network meta-analyses (NMA). ***Methods:*** Standard parametric, mixture- (MCM) and non-mixture-cure (nMCM), piecewise, and splines are used for comparison. For the base case model, selection treatment coefficients were assigned to all parameters. The lowest leave-one-out information criterion (LOOIC) defined the base case parametric distribution per flexible model. The best-fitting distribution per model was rerun with treatment coefficients only for those larger than their standard deviation, and the one with the lowest LOOIC was considered the base case. Models were compared based on incremental mean survival and uncertainty. We used a network of four trials in previously treated advanced non-small cell lung cancer, comparing nivolumab vs. docetaxel, pembrolizumab vs. docetaxel, and two trials comparing atezolizumab vs. docetaxel. ***Results:*** The base case parametric distributions were log-logistic for standard parametric distribution, log-logistic for MCM, lognormal for nMCM, Weibull for mixture, log-logistic for piecewise, and log-logistic for spline models, with corresponding LOOICs 13418, 13417, 13420, 13417, 13418, and 13419. Regarding treatment effect parametrization, we applied no treatment effect on cure, for instance, but treatment effects on the scale for MCM and nMCMs. For piecewise, treatment effects were only applied on the second piece. After reducing the number of treatment effects, the mixture Weibull had the lowest LOOIC with 13407. The mean incremental survival with docetaxel as reference for mixture Weibull was 3.58 [95%CI 1.2,5.26], 0.45[0.19,0.93] and 2.67[0.2,9.48] for nivolumab, pembrolizumab and atezolizumab, respectively. For standard log-logistic these were 1.67 [0.94,2.55],1.01[0.55,1.61], and 0.43[0.17,0.77]. ***Conclusions:*** Treatment effect specification is important as outcomes and uncertainty differ over the tested models.

Methodological Developments in Survival Analytic Methods: to Inform Cost-Effectiveness Models

P45
THE USE OF HISTORICAL CLINICAL TRIAL DATA TO INFORM SURVIVAL EXTRAPOLATION

Pham HA,[1] Smalbrugge D,[2] Kroi F,[2] Heeg B,[1] Ouwens M[3]
[1]*Ingress-health, Rotterdam, ZH, Netherlands,* [2]*Ingress-health, Rotterdam, Netherlands,* [3]*AstraZeneca, Mölndal, O, Sweden*

Objectives: Standard parametric distributions are commonly used for the extrapolation of survival data in cost-effectiveness analyses. However, survival data is often immature and uncertainty remains around the survival extrapolations. Mature historical data can be used to better predict survival beyond trial data. This study assessed two methods to incorporate historical data in the extrapolation of immature survival data. ***Methods:*** Immature data of a breast cancer trial comparing pertuzumab+trastuzumab+docetaxel versus trastuzumab+docetaxel (follow-up time 38 months; data-cut 2015) was extrapolated and mature survival data (follow-up time 120 months; data-cut 2020) from the same trial was used to validate the extrapolations. The historical data was from a previous breast cancer trial including mature survival data of trastuzumab+docetaxel (follow-up time 50 months; data-cut 2005). Two methods to quantitatively inform the extrapolation of immature survival data with historical data were compared to standard parametric distributions: 1) historical shape parameter as informative prior for the shape of the immature data; 2) historical data as a third arm. Predictions were assessed with delta area under the curve (AUC) values based on the mature survival data. ***Results:*** Without priors, the delta AUC was 7.59, 1.62, 13.15, 8.32, 25.15, and with the historical arm the delta AUCs were 9.65, 4.38, 6.79, 8.26, 21.81, for Weibull, loglogistic, lognormal, exponential, and Gompertz, respectively. With priors, the delta AUC were 8.43, 3.37, 9.11, 23.68, for Weibull, loglogistic, lognormal, and Gompertz, respectively (as for exponential there is no shape parameter). The loglogistic distribution without priors predicted the immature data the best. For three out of five distributions, the extrapolations with a historical arm resulted in better predictions compared to the extrapolations without prior. ***Conclusions:*** The impact of external data on clinically plausible survival extrapolations can further be improved by using historical data with longer follow-up with treatment patterns similar to the current standard of care.

P46
EFFECTIVE USE OF RECONSTRUCTED SURVIVAL AND COMPARATIVE EFFECTIVENESS DATA: A CASE STUDY FROM ESTIMATING UNREPORTED SUBGROUP SURVIVAL IN ADVANCED STAGE GASTROINTESTINAL CANCERS

Alagoz O,[1] Xiao H,[2] Singh P,[3] Gricar J,[3] Dixon M,[3] Kim I,[3] Kurt M[2]
[1]*University of Wisconsin-Madison, Madison, WI, USA,* [2]*Bristol Myers Squibb, Princeton, NJ, USA,* [3]*Bristol Myers Squibb, Lawrenceville, NJ, USA*

Objectives: This study devises a systematic approach that can utilize aggregate level survival and comparative effectiveness data published from randomized controlled trials (RCT) to assist subgroup-specific health economic and meta-analyses. ***Methods:*** We developed a soft-constrained optimization model, which approximates the restricted mean survival time (RMST) for the overall population in each arm via weighted sum of the RMSTs of two subgroups of interest. Survivals of both subgroups in each arm were assumed to follow Weibull or log-logistic distribution. The constraint ensured that cumulative hazards between the arms were proportional for each subgroup at a sufficiently long pre-specified time point. Estimated subgroup-specific survival functions for the control arm were direct outputs of the model and were shifted by applying the reported hazard ratios from the forest plots to generate their counterparts for the intervention arm assuming proportional hazards between the arms. For validation, we tested our approach in a case study consisting of 10 distinct RCTs with reported subgroup-specific Kaplan-Meier (KM) curves from advanced stage gastrointestinal tumors. ***Results:*** Across all 48 subgroups, on average, loglogistic model performed equally or better than Weibull model in performance criteria comparing overall survival (OS) rates, median OS and RMSTs. Predicted survival curves laid within the 95% confidence intervals (CIs) of reported KM-curves in 75% and 81% of the time for Weibull and loglogistic models, respectively. Predicted median survivals were within the 95% CIs of the reported medians in 34 and 40 subgroups for Weibull and loglogistic models, respectively. Average relative gap between the predicted and reported RMSTs was 10% in both models. Predicted RMSTs were within the 95% CI of reported RMSTs in 34 and 37 subgroups for Weibull and loglogistic models, respectively. ***Conclusions:*** Our elicitation approach is effective and demonstrably reliable in deriving unreported subgroup survival with flexible time-varying hazard functions.

P47
A BAYESIAN HIERARCHICAL MIXTURE CURE MODELLING (MCM) FRAMEWORK FOR THE JOINT UTILIZATION OF PROGRESSION FREE SURVIVAL (PFS) AND OVERALL SURVIVAL (OS) IN ESTIMATING LONG-TERM SURVIVORSHIP RATES IN PREVIOUSLY UNTREATED METASTATIC MELANOMA: A CASE STUDY FROM CHECKMATE-067 TRIAL

Green N,[1] Paly V,[2] Youn JH,[3] Kurt M,[4] Moshyk A,[4] Gianluca Baio G[1]
[1]*University College London, London, LON, UK,* [2]*ICON plc, New York, NY, USA,* [3]*ICON plc, Marlow, Bucks, UK,* [4]*Bristol Myers Squibb, Princeton, NJ, USA*

Objectives: Differences in emergent survival plateaus between PFS and OS may imply clinically unintuitive dichotomy between the resulting proportions of long-term survivors (LTS) when they are analyzed separately via mixture cure models (MCM). We present a novel Bayesian hierarchical (BH) MCM framework assuming a dependency between PFS and OS to estimate LTS rates in CheckMate 067 and demonstrate its practical utility over frequentist MCMs in long-term QALY estimations. ***Methods:*** Frequentist and BH MCMs were fitted to PFS and OS data from the trial with minimum 60-months of follow-up. In the frequentist MCMs, PFS and OS were modelled separately whereas in BH MCMs both endpoints were modelled jointly with a shared LTS rate. In both approaches, background mortality rates were taken from World Health Organisation's age, gender and country-specific lifetables and time-to-event outcomes for the non-LTS were modeled using a range of standard parametric distributions. Estimated incremental QALYs gains for nivolumab containing therapies versus ipilimumab under both approaches were compared using

local tariffs from US, Canada, UK, France, Sweden, Belgium, Netherlands, Portugal and Australia. ***Results:*** Among all combinations of distributions considered for the BH MCM, the exponential-exponential fit adequately captured the observed survival trends for both endpoints with reasonable goodness-of-fit measures and shared LTS rates which were (95% credible intervals) 46.3% (32.8%, 62.6%) for nivolumab+ipilimumab, 37.8% (21.6%, 55.7%) for nivolumab, and 15.1% (6.8%, 26.0%) for ipilimumab. For each arm, shared LTS-rates were in between individually-estimated LTS rates from the OS and PFS data in the frequentist MCMs. Compared to frequentist MCMs, over 20-years BH MCMs produced higher incremental QALY gains for nivolumab+ipilimumab and nivolumab versus ipilimumab with differences ranging from 0.30-0.48 and 0.17-0.26, respectively. ***Conclusions:*** Our BH MCM framework can alleviate the disparity between individually estimated OS- and PFS-based LTS rates, and allow for more robust clinical inference and extrapolations.

P48
JOINT MODELLING MEASURABLE TARGET LESIONS, EQ5D/UTILITY AND OVERALL SURVIVAL: DO WE STILL NEED PARTITIONED SURVIVAL MODELS?

Felizzi F,[1] Upton A,[2] **Paracha N**[3]
[1]*Novartis, Basel, BS, Switzerland,* [2]*Bayer, Whippany, NJ, USA,* [3]*Bayer, Basel, Switzerland*

Objectives: The biologic effect of treatment is typically analyzed using tumor burden or PFS, the latter often leads to loss of information due to being based on the categorization of RECIST criteria. Whereas, tumor burden is complete and rational summarization of the process which captures data on the primary mechanism through which most treatments are expected to act. The objective of this study is to estimate overall survival using joint model which simultaneously models patient's longitudinal tumor burden (TB) trajectory, EQ-5D/utility and survival by accounting for the association between the three outcomes. ***Methods:*** We simulated data based on a clinical trial (n=110), repeated measure (940 observations). Tumor burden is defined as sum of the longest diameters of measurable prespecified target lesions. Joint model using current value association structure and several distribution was fitted to predict survival. Time was included as a random effect. No covariates were included in the model. We compared joint model with independently fitted models for OS and PFS. ***Results:*** The median follow-up visit was 25 months (range: 22-29 months). Joint model showed a positive statistical association to TB, indicating that a higher value of TB increases the risk of death and a negative statistical association to EQ-5D/utility, indicating that proximity to death is linked to lower utility values. While the restricted mean estimates did not show differences between independently fitted models and the joint model, the overall uncertainty around long-term outcomes is dramatically reduced. ***Conclusions:*** Multivariate Joint modelling is a powerful tool which can simultaneously model multiple longitudinal biomarkers and survival that are relevant in economic evaluation and leads to more efficient treatment effect. Given the limitations associated with partitioned survival model, i.e. the structural assumptions that survival functions modeled are independent, this approach provides a new way to conduct cost-effectiveness analysis for metastatic oncology treatments.

Methods and Controversies in the Evaluation of Oncology Products

P49
ACCURACY OF LIFE YEAR GAINS PREDICTIONS FOR CAR-T THERAPY IN THE LONG TERM: AN ANALYSIS FOR AXICABTAGENE CILOLEUCEL IN REFRACTORY LARGE B-CELL LYMPHOMA

Porteous A,[1] Gregori D,[1] Hilton B[2]
[1]*Costello Medical, London, UK,* [2]*Costello Medical, Cambridge, UK*

Objectives: Survival profiles for chimeric antigen receptor T cell (CAR-T) therapies commonly exhibit a plateau that may indicate a cure. This leads to challenges predicting long-term outcomes for novel CAR-T therapies when trial data are immature. This study retrospectively analysed the accuracy of overall survival (OS) extrapolations from interim data cuts in predicting realised long-term life years (LYs) for axicabtagene ciloleucel in patients with refractory large B-cell lymphoma. ***Methods:*** Published OS data for axicabtagene ciloleucel from successive data cuts of ZUMA-1 (median follow-ups of 15.4, 27.1 and 51.1 months) were digitised. Standard parametric, spline (1–2 knots; normal, odds and hazard) and mixture cure models (MCMs) were fitted to the first two data cuts. Statistical fit was tested using Akaike and Bayesian information criteria (AIC and BIC). Cumulative LYs were estimated for each model over a 58-month time horizon, corresponding to the longest duration of published OS data. These projected LYs were then compared to realised LYs over this period. ***Results:*** At the earliest data cut, MCMs provided the best predictions of realised LYs, with a mean absolute difference of 6.8% between predicted and realised LYs across MCM models (range: 0.5%–10.7%). Standard parametric extrapolations considerably underestimated realised LYs (mean absolute difference: 19.2%; range: 9.9%–28.0%), whilst spline models offered a mean absolute difference of 8.3% (range: 1.4%–13.2%). Similar findings were observed for extrapolations based on the second data cut (representing 11.7 months additional follow-up), but differences between model classes were less pronounced. ***Conclusions:*** MCMs may offer the best predictions of long-term survival for CAR-T therapies, particularly when only short-term data are available. Standard parametric models may be inappropriate to predict survival when extrapolating immature data, failing to capture the plateau in survival typical of CAR-T therapies. Further research is required to determine whether these findings are generalisable across CAR-T therapies and indications.

P50
COMPETING RISK AND MULTISTATE MODEL COMPARED TO PARTITIONED SURVIVAL MODEL IN METASTATIC NON-SMALL CELL LUNG CANCER

Le Mezo A,[1] Kandel M,[2] Caillon M,[3] Chauny JV,[4] Borget I[5]
[1]*Master 2 Market-Access et Evaluation Médico-Economique, Université Paris-Saclay, Sèvres, France,* [2]*IQVIA France, La Défense Cedex, France,* [3]*Amgen SAS, Boulogne Billancourt, 75, France,* [4]*Amgen SAS, Boulogne Billancourt, France,* [5]*Department of Biostatistics and Epidemiology, Gustave Roussy, Paris-Saclay University, Villejuif, France; Oncostat; GRADES, Paris-Saclay University, Châtenay-Malabry, France U1018, Inserm, Paris-Saclay University, "Ligue Contre le Cancer" labeled team, Chatenay-Malabry, 92, France*

Objective: Partitioned survival analysis (PartSA) is commonly used for economic evaluations in oncology. We compare PartSA to a competing risk semi-markov multistate model (MSM) and investigate differences in estimated cost-effectiveness in metastatic non-small cell lung cancer from a French perspective. ***Methods:*** To populate both models, pooled data from Checkmate 017 and 057 was digitized to reconstruct patient-level data from overall survival, progression-free survival and post-progression survival Kaplan Meier curves. Models consisted of three states: progression-free, progressed disease and death. Statistical analysis was performed under *R* to fit parametric survival models for the PartSA and to estimate state transitions for the MSM (package "mstate"). Costs and utilities were integrated in the model using values from nivolumab's French HTA submission. Outcomes were discounted at 2,5% per annum over a 7-year time horizon. ***Results:*** Both models produced similar results during trial period (15 months) but showed discrepancies during extrapolation, inducing different costs and quality-adjusted life years (QALYs) over time. Incremental life years gained (LYG) of nivolumab versus docetaxel were 1,11 LYG (PartSA) and 0,88 LYG (MSM). However, both incremental cost effectiveness ratios (ICER) were similar, 88 979€/QALY (PartSA) and 90 148€/QALY (MSM). Scenarios assessing different parametric extrapolations produced an average of 112 496€/QALY (PartSA) and 113 085€/QALY (MSM) with coefficients of variation (CV) estimated at 19,0% (PartSA) and 16,3% (MSM). Probabilistic ICER results were similar between models with CV estimated at 21,8% (PartSA) and 23,2%(MSM). ***Conclusions:*** This study provides new evidence on structural uncertainty. The MSM produced similar results to the PartSA for a 15 months follow-up. Ideally, both models should be tested to validate extrapolations and to ensure that the structural uncertainty is limited. However, when the time horizon is short (<10 years), the impact of the model choice on the ICER seems limited.

P51
BLENDED SURVIVAL CURVES: A NEW APPROACH TO EXTRAPOLATION FOR TIME-TO-EVENT CLINICAL TRIAL DATA IN HEALTH TECHNOLOGY ASSESSMENT

Che Z,[1] Baio G,[2] Green N[2]
[1]*University College London, London, Great Britain,* [2]*University College London, London, LON, UK*

Objectives: Survival extrapolation is generally required in the cost-effectiveness analysis to estimate the survival benefit of a new intervention, due to the limited duration of randomized controlled trials (RCTs). Current techniques of extrapolation often assume constant treatment effect beyond the observed period in the RCT, which is implausible and highly influential in survival estimates for resource allocation decisions. The objective of this study is to develop a novel methodology based on "blending" survival curves as a possible solution. ***Methods:*** We mixed a flexible Cox semi-parametric model conducted in Bayesian setting to fit the observed data and a parametric model either by prior assumptions or external data on the long-term expected behavior of the underlying survival curves. The two are "blended" into a single survival curve that is equivalent to the Cox model over the follow-up intervals and gradually approaching to the parametric model over the extrapolation period based on a weight function. The weight function and mixing area of the blended curve control the way the internal and external data sources influence the estimated survival. ***Results:*** A 4-year follow-up RCT of rituximab in combination with fludarabine and cyclophosphamide (RFC) v. fludarabine and cyclophosphamide alone (FC) for the first-line treatment of chronic lymphocytic leukemia is used to illustrate the method. Two kinds of prior information, registry data and summary of clinical knowledge were respectively used for the long-term estimate. ***Conclusions:*** Long-term extrapolation with various assumptions of treatment effect may give significantly different estimated mean survival gains. The blending process allows a consideration of plausible scenarios, abandoning the over-optimistic constant treatment effect and provides sufficient flexibility. Not only internal but also external validity could be carefully considered since a wide range of external evidence can be used to inform the long-term estimate, including hard data from real world and clinical expert opinion.