ASSESSING LONG-TERM BENEFITS OF IMMUNOTHERAPY BASED ON EARLY TUMOR ASSESSMENT DATA

PRALAY MUKHOPADHYAY

AstraZeneca, Gaithersburg, USA

XUEKUI ZHANG

University of Victoria, Victoria, Canada

Disclosures

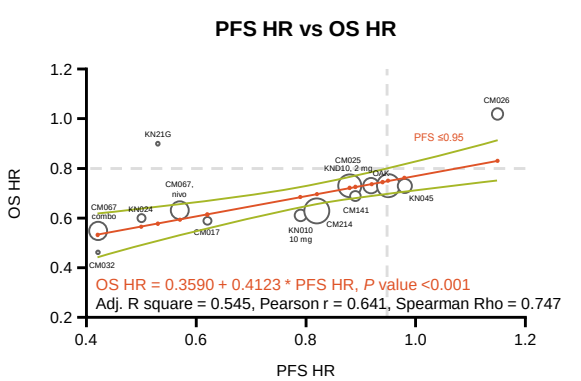
- Pralay Mukhopadhyay is an employee of AstraZeneca. The views and opinions expressed herein are solely his own and should not necessarily be construed to represent those of AstraZeneca or its affiliates.

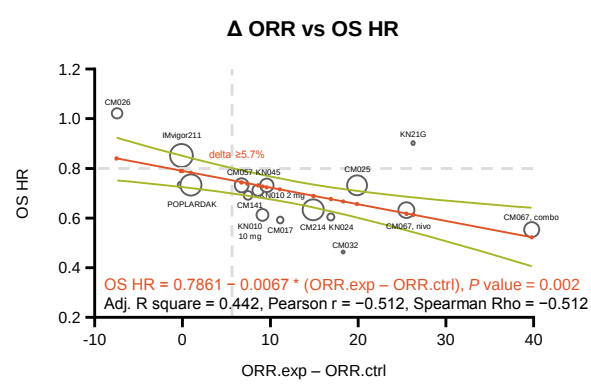
Background

- The emerging data in immuno-oncology have changed the way many cancers are being treated
- Clinically meaningful benefits in OS have been observed in many tumor types, including NSCLC, melanoma, renal cell carcinoma and SCCHN
- OS remains the gold standard of clinical benefit in patients with cancer
- However, challenges remain with detecting OS benefits, including duration of follow-up, crossover and competing risks (in earlier disease settings)
- PFS and ORR are often not reliable surrogates for predicting OS
- What novel approaches can be considered for making an early assessment?

NSCLC, non-small cell lung cancer; ORR, objective response rate; OS, overall survival; PFS, progression-free survival; SCCHN, squamous cell cancer of the head and neck

Modest correlation of PFS HR vs OS HR and Δ ORR vs OS HR in patients treated with I-O vs chemotherapy in various solid tumors





Modest improvements in PFS may result in a large improvement in OS

Modest improvements in ORR may result in a large improvement in OS

A PFS HR ≤ 0.9 or Δ ORR > 10% may be associated with an OS HR of ≤ 0.8

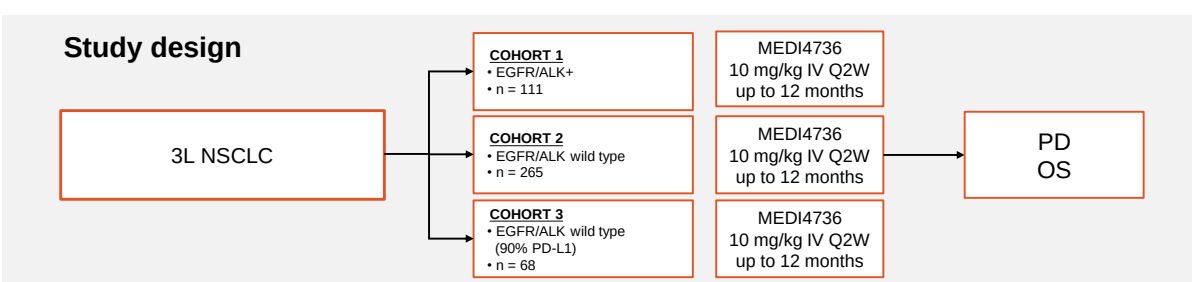
Adj, adjusted; CM, CheckMate; ctrl, control; exp, experimental; HR, hazard ratio; I-O, immuno-oncology; KN, Keynote; ORR, objective response rate; nivo, nivolumab; OS, overall survival; PFS, progression-free survival; SCCHN, squamous cell cancer of the head and neck

Tumor kinetics modeling: an alternative approach?

- Using baseline characteristics and early tumor assessment data, we calculated a score, chose a cutoff value, and used these values to segment patients into groups
- To build this scoring system, we:
  - used data from a study in late-stage NSCLC (ATLANTIC)
  - trained a model to predict best overall response (PR/CR) to treatment
  - obtained a formula to calculate predicted probability of response, and used it as the score
- We evaluated OS difference between segmented groups to check performance of our rule
- We then used a second study (study 1108) to independently validate the rule built on ATLANTIC, to show that our proposed algorithm can be used for prediction in future studies

CR, complete response; NSCLC, non-small cell lung cancer; OS, overall survival; PFS, progression-free survival;
PR, partial response; SCCHN, squamous cell cancer of the head and neck

ATLANTIC: phase 2, 3L NSCLC durvalumab monotherapy



- Primary endpoint: ORR
- Follow-up for OS
- First patient in: Q1 2014
- Each arm analyzed separately
- Recruitment into cohorts 1 to 3 closed in Q4 2015

3L ,third line; ALK, anaplastic lymphoma kinase; EGFR, epidermal growth factor receptor;
IV, intravenous; NSCLC, non-small cell lung cancer; ORR, objective response rate;
OS, overall survival; PD, progressive disease; PD-L1, programmed death ligand 1;
Q1, first quarter; Q2W, every 2 weeks; Q4, fourth quarter

Segmentation model

- To segment patients, we built a model to predict BOR (CR/PR vs others)

  BOR (0/1) ≈ baseline characteristics + early tumor assessment results

- Baseline characteristics included the following
  - Region
  - Sex
  - Age
  - Race
  - PD-L1 status
  - Histology
  - Smoking status
  - Performance status
  - Cancer stage
  - Line of therapy
  - Tumor location
  - Baseline tumor size

- Early tumor assessments
  - We used percent change from baseline at the first two or three follow-up tumor assessments

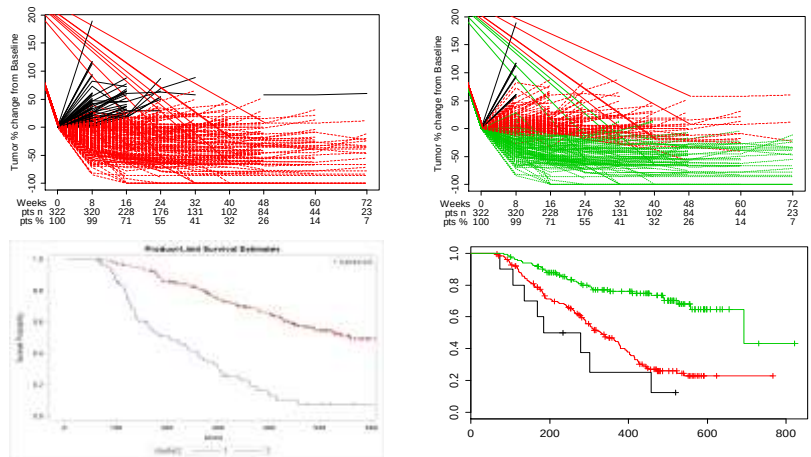
BOR, best overall response; CR, complete response; PD-L1, programmed death ligand 1; PR, partial response

Tumor kinetics model

- There are multiple approaches to modeling tumor kinetics
- We considered/compared the following eight approaches
  - Model 1: first principal component score (> 90% of variance)
  - Model 2: tumor size percent change of first 2 or 3 follow-up visits
  - Model 3: 2-cluster membership
  - Model 4: 3-cluster membership
  - Model 5: 4-cluster membership
  - Model 6: 5-cluster membership
  - Model 7: deterministic percent rule (group patients by “≥ 2 visits with ≥ 10% tumor size reduction”)
  - Model 8: no on-treatment tumor information (to evaluate added value of tumor kinetics)

Example of clustering tumor growth profile

- A K-means clustering algorithm based on longitudinal tumor assessment was used to group patients into two or three clusters



pts = patients

The best model?

- Misclassification* rate was used to evaluate the performance of models

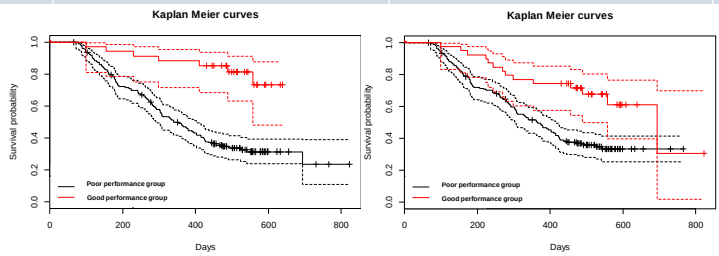
	PCA	% change in tumor	Two clusters	Three clusters	Four clusters	Five clusters	Fixed % rule	No tumor
Using three visits	7.85%	8.38%	16.75%	16.75%	6.28%	6.81%	20.42%	16.75%
Using two visits	9.42%	9.42%	16.75%	19.37%	8.90%	9.95%	19.37%	16.75%

  - Using two visits did not result in much worse performance than using three visits, but because we could predict results much earlier, we chose two visits
  - The 4-cluster model had the best performance, but this clustering rule is harder to transfer to other studies than other models
  - The PCA model is not much worse than the best (4-cluster) model, therefore we decided to use this approach
    - The first PC can be interpreted as the weighted average of tumor percent reductions
  - No on-treatment tumor information and the fixed-rule models have the highest misclassification rates

*Misclassification refers to predicting true responders as non-responders and vice-versa
PC, principal component; PCA, principal component analysis

Comparing observed survival outcomes of predicted groups

	Proposed method		Fixed rule		No tumor	
	Poor performance group (n = 157)	Good performance group (n = 34)	Poor performance group (n = 152)	Good performance group (n = 39)	Poor performance group (n = 191)	Good performance group (n = 0)
Nonresponders	149	10	137	22	159	0
Responders	8	24	15	17	32	0
Median survival (95% CI), days	340 (292–403)	NA (557–NA)	294 (237–409)	769 (682, NA)	410 (340–490)	NA
6-month survival (95% CI), days	0.742 (0.665–0.804)	0.941 (0.785–0.985)	0.740 (0.662–0.803)	0.923 (0.780, 0.975)	0.778 (0.711–0.831)	NA
1-year survival (95% CI), days	0.478 (0.397–0.554)	0.882 (0.716–0.954)	0.501 (0.418–0.578)	0.744 (0.576, 0.853)	0.552 (0.477–0.620)	NA
PH ratio (95% CI) / *P* value	0.2059 (0.0569–0.4437) / 0.00005		0.4097 (0.2334–0.719) / 0.00187		NA	



Good-performance vs poor-performance groups were based on their cross-validation prediction of responders

CI, confidence interval; NA, not applicable; PH, prediction hazard

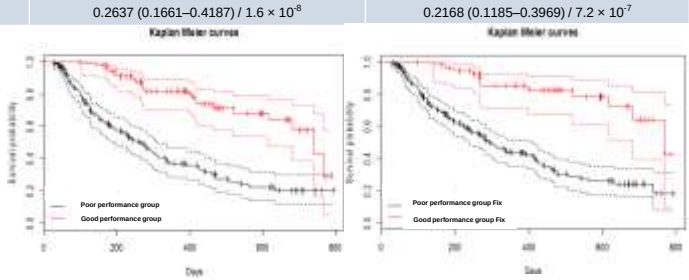
Patient-segmentation rule

- We refitted the PCA model to full ATLANTIC data to generate a patient-segmentation rule
  - This rule will be used to segment patients in future studies
- The selected variables were:
  - Sex: 0.028 (male response better)
  - Histology group: 0.188 (squamous response better)
  - Smoker group: 0.034 (smoker response better)
  - Line of therapy: −0.176 (earlier-line response better)
  - First PC of tumor percent change from baseline: −0.041 (larger reduction response better)
- Linear combination of the above variables was converted into a probability of response (ie, CR/PR) via a logit link
- A cutoff probability of 0.278 was selected through the second layer of cross-validation and used to segment patients into two groups

CR, complete response; PC, principal component; PCA, principal component analysis; PR, partial response

Validated patient-segmentation rule in study 1108

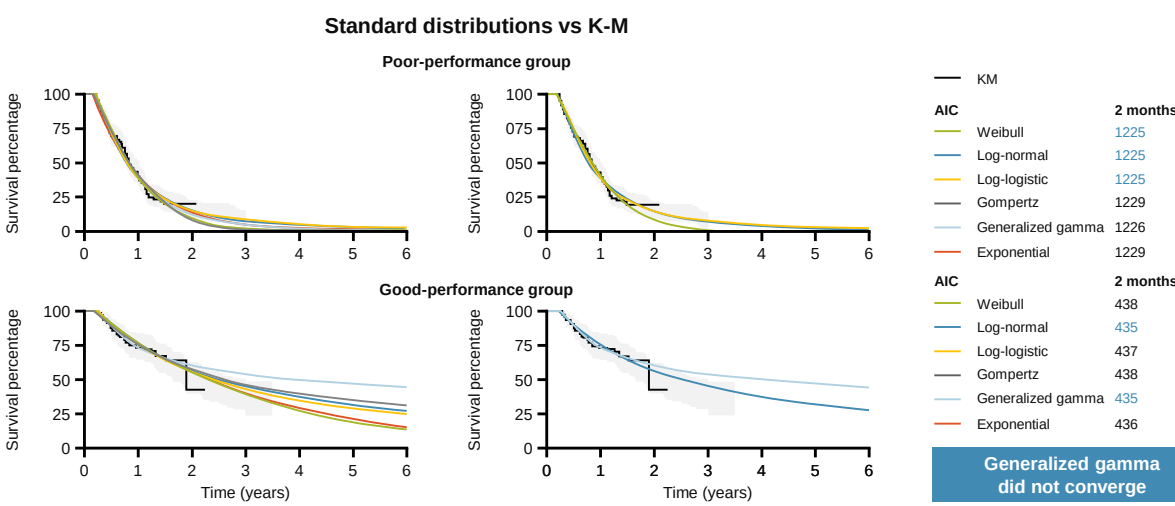
	Proposed method		Fixed rule		No tumor	
	Poor performance group (n = 154)	Good performance group (n = 82)	Poor performance group (n = 182)	Good performance group (n = 54)	Poor performance group (n = 236)	Good performance group (n = 0)
Nonresponders	154	34	173	15	188	0
Responders	0	48	9	39	48	0
Median survival (95% CI), days	265 (194–315)	739 (616–NA)	294 (237–409)	769 (682–NA)	425 (310–519)	NA
6-month survival (95% CI), days	0.605 (0.517–0.682)	0.937 (0.855–0.973)	0.653 (0.574–0.720)	0.962 (0.856–0.990)	0.729 (0.664–0.783)	NA
1-year survival (95% CI), days	0.371 (0.282–0.460)	0.819 (0.708–0.891)	0.434 (0.349–0.516)	0.853 (0.715–0.927)	0.542 (0.468–0.611)	NA
PH ratio (95% CI) / *P* value	0.2637 (0.1661–0.4187) / 1.6×10^{-8}		0.2168 (0.1185–0.3969) / 7.2×10^{-7}		NA	



Good-performance vs poor- performance groups were predicted by rule built on the ATLANTIC study

CI, confidence interval; NA, not applicable; PH, prediction hazard

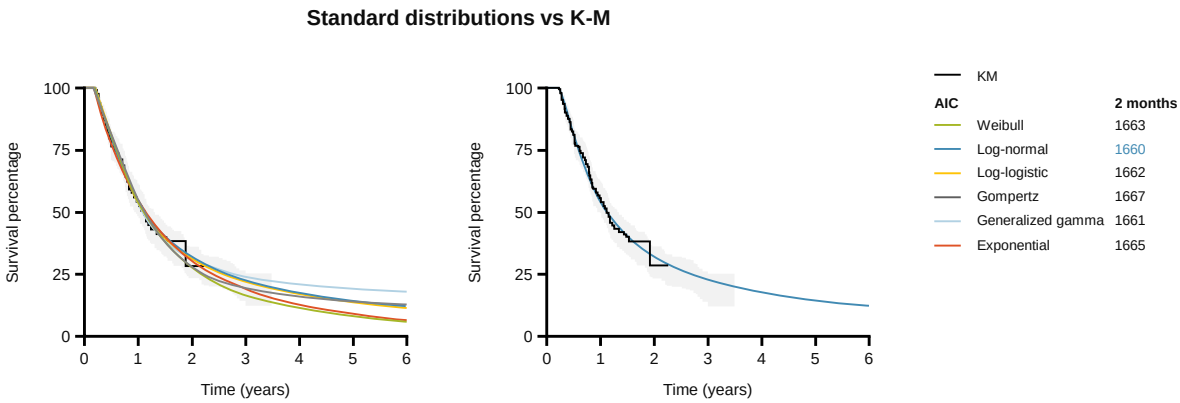
OS extrapolation using predictive modeling from the ATLANTIC study



Generalized gamma did not converge

AIC, Akaike information criterion; K-M, Kaplan-Meier; OS, overall survival

OS extrapolation using predictive modeling in the overall population



AIC, Akaike information criterion; K-M, Kaplan-Meier; OS, overall survival

Conclusions

- We built a rule to segment patients into two groups (predicted responders vs nonresponders), using baseline characteristics and early tumor measurements
- We validated our segmentation rule in an independent study
  - The rule shows 100% sensitivity (it detected all true responders)
  - Survival outcomes were significantly different between the two segmented groups
- We hope to extend this approach using data from RCTs to identify predictive models that may help to differentiate between patients receiving I-O and 'non' I-O therapies
- The above approach could also be used in the context of OS extrapolation, instead of using traditional methods to segment patients, such as with landmark models
  - Modeling OS separately in predicted responders and non-responders, allows us to take into consideration the heterogeneity of these populations based on their underlying disease trajectory

I-O, immuno-oncology; RCT, randomized controlled trial; OS, overall survival

Acknowledgments

- Mario Ouwens, PhD, AstraZeneca, Gothenburg, Sweden
- Jiabu Ye, PhD, AstraZeneca, Gaithersburg, USA