Evaluation of surrogate endpoints for previously untreated unresectable or metastatic melanoma: Analyses from a longitudinal electronic health record database in the United States

Leung L,[1] Mohr P,[2] Serafini P,[1] Kanters S,[1] Pourrahmat MM,[1] Moshyk A,[3] Srinivasan S,[3] Kurt M[3]

[1]Evidinno Outcomes Research Inc., Vancouver, BC, Canada; [2]Elbe Klinikum Buxtehude, Buxtehude, NI, Germany; [3]Bristol Myers Squibb, Lawrenceville, NJ, US

Presentation number P49

Disclosures

- This study was conducted by Evidinno Outcomes Research Inc. in collaboration with Bristol Myers Squibb, which provided financial sponsorship
- Patient-level real-world data (RWD) were shared as per the mutual agreement between Bristol Myers Squibb and Flatiron Health

Background: Melanoma

- An estimated 97,610 new cases of melanoma of the skin are expected in 2023 within the United States (US); 7,990 patients are estimated to die of the disease[1]
- Melanoma is easily treated if caught early, but melanomas that form on the back or neck may not be detected until the cancer has metastasized, at which point the historic 5-year survival rate is low (32%).[1] As a result, there is a growing need for new treatments to improve overall survival (OS)
- The advent of modern treatments such as immune checkpoint and BRAF-MEK inhibitors has revolutionized the treatment of advanced melanoma, for which median survival can be from few to 5+ years depending on the treatment regimen
- Most pivotal trials of these modern treatments to date have used OS as a (co-) primary endpoint;[2-5] however, OS is the lengthiest time-to-event outcome to collect with statistical maturity
- Using intermediate endpoints that may reach statistical maturity sooner could expedite drug development and improve patient access to novel treatments

1. Siegel RL, et al. *CA Cancer J Clin* 2023;73(1):17-48. 2. Wolchok J, et al. *J Clin Oncol* 2022;40(2):127-137. 3. Tawbi HA, et al. *N Engl J Med* 2022;386(1):24-34. 4. Robert C, et al. *Lancet Oncol* 2019;20(9):1239-1251. 5. Robert C, et al. *N Engl J Med* 2019;381(7):626-636.

Background: Surrogate endpoints in advanced melanoma

- Most prior research on surrogate endpoints (SEs) for OS in advanced melanoma has been mostly restricted to analysis of individual- or aggregate-level randomized controlled trial (RCT) data or specific therapies based on their mechanism of action:
  - Flaherty et al. (2014)[6] assessed whether progression-free survival (PFS) can be regarded as an SE for OS through a meta-analysis of RCTs that had dacarbazine as the control group and any systemic therapy as the experimental arm
  - Leung et al. (2022)[7] and Nie et al. (2020)[8] investigated PFS as an SE using aggregate-level RCT data identified via systematic reviews. The latter study, Nie et al. (2020),[8] focused on immune checkpoint inhibitor trials only
  - Branchoux et al. (2022)[9] explored the composite endpoint of time-to-next-treatment or death (TNTD) as an SE using patient-level data from CheckMate 067
  - Larkin et al. (2022)[10] studied PFS, TNTD, and complete- and objective-response rates as SEs using pooled patient-level data from four different CheckMate trials (-066, -067, -069, and -511)
  - Mohr et al. (2022)[11] investigated PFS, TNTD, and time-to-treatment discontinuation (TTD) as SEs using RWD from patients treated with pembrolizumab in a German registry

6. Flaherty KT, et al. *Lancet Oncol* 2014;15(3):297–304. 7. Leung L, et al. *Value Health* 2022;25(1):S22. 8. Nie RC, et al. *Ther Adv Med Oncol* 2020;12:1-10. 9. Branchoux S, et al. *ESMO Open* 2022;7(1):100340. 10. Larkin J, et al. *Ann Oncol* 2022; 33(7):S919-S920. 11. Mohr P, et al. *Cancers* 2022;14(7):1804.

Background: Surrogacy

- An SE can be validated in place of OS to assess the efficacy of a new therapy earlier
- Elston et al. (2009)[12] identify three levels of evidence needed for a surrogacy relationship:
  1. The endpoints' treatment effects are correlated at the aggregate level
  2. The endpoints are correlated at the individual patient level
  3. A relationship between the endpoints is biologically plausible
- The "two-level meta-analytic approach" based on individual-level patient data from RCT settings is currently the most preferred and robust statistical process in the literature for the validation of SEs:[13]
  - **Individual-level** association seeks to address the prognostic role of the SE for OS
  - **Treatment-effect** association seeks to address if the OS benefit can be predicted from the benefit on the SE
- To date, only Mohr et al. (2022)[11] has investigated SEs in melanoma using RWD. In this study, due to lack of variability in treatment, no investigations were made for correlations between treatment-effects on SEs and OS

11. Mohr P, et al. *Cancers* 2022;14(7):1804. 12. Elston J, et al. *Int J Technol Assess Health Care* 2009;25(1):6-13. 13. Burzykowski T, et al. *The Evaluation of Surrogate Endpoints.* New York, Springer, 2005.

Objectives

- To investigate PFS, TNTD, TTD and time-to-progression (TTP) as candidate SEs for OS in previously untreated unresectable or metastatic melanoma (mMel), using RWD
- As secondary objectives with no particular order, to:
  - Assess the sensitivity of PFS-OS correlation with respect to alternate descriptions of PFS around the designation of subsequent treatment initiations
  - Identify a ranking of the strength of candidate SEs in their association with OS
  - Evaluate the utility and validity of the model both internally and externally
  - Compare findings to prior research applied on data from RCTs and RWD

Methods: Study population & patient eligibility criteria

- The US Flatiron Health database, which is a retrospective and longitudinal database comprising de-identified RWD derived from electronic health records from 265 US cancer clinics was used
- The database included 11,891 patients diagnosed with advanced melanoma between 1968 and 2021
- Patients in the initial cohort were eligible for surrogacy assessments if they:
  - Had an initial diagnosis of unresectable or mMel (IIIB+; Flatiron does not indicate the American Joint Committee on Cancer [AJCC] edition)
  - Initiated the first-line (1L) therapy on or after the initial diagnosis for advanced melanoma between January 1, 2011 and September 30, 2020 (allowing time for follow-up until August 31, 2021)
  - Were at least 18 years old at diagnosis
  - Had at least one month of pre- and post-1L initiation data
  - Had complete covariate data (age, race, sex, Eastern Cooperative Oncology Group [ECOG], and BRAF mutation status)
  - Were not part of a clinical trial
  - Had complete ICD diagnosis data and no primary or secondary malignancy pre-1L treatment

Methods: Endpoint definitions

- All SEs and OS are measured from the first dose of 1L treatment

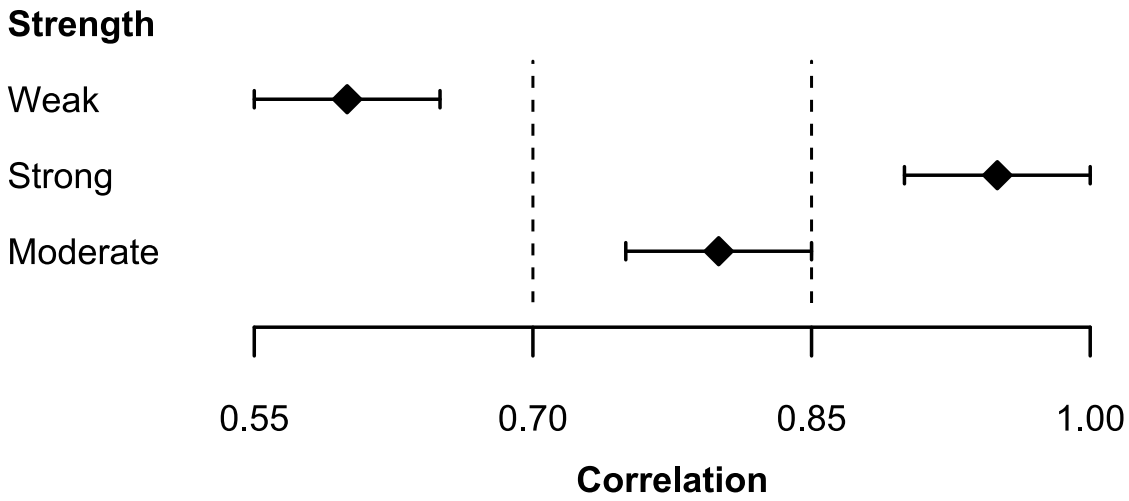
Endpoint	Event(s)	Censor event(s)
OS	All-cause mortality	Last electronic health record (EHR) activity
PFS[a]	Disease progression All-cause mortality	2L treatment initiation Last EHR activity
TNTD	2L treatment initiation All-cause mortality	Last EHR activity
TTD	Last dose of 1L treatment	At the last dose of 1L treatment if 1) No initiation of 2L treatment 2) No visit activity more than 120 days after the last dose 3) Alive at the last dose
TTP	Disease progression	2L treatment initiation All-cause mortality Last EHR activity

[a]A secondary definition of PFS in which 2L treatment initiation was defined as an event was also assessed.
1L, first-line; 2L, second-line; OS, overall survival; PFS, progression-free survival; TTD, time-to-treatment discontinuation; TTP, time-to-progression, TNTD, time-to-next-treatment or death.

Methods: Individual-level surrogacy

- Individual-level association between each SE and OS was assessed in a non-parametric fashion using Spearman's rank correlation (ρ) coefficient, with a 95% confidence interval (CI) estimated using bootstrapping
- Strength of association between each SE and OS was evaluated using criteria adapted from the Institute for Quality and Efficiency in Health Care (IQWiG) by Kemp et al. (2017):[14]
  - Weak: upper limit of 95% CI of ρ ≤ 0.70
  - Strong: lower limit of 95% CI of ρ ≥ 0.85
  - Moderate: anything else

An example demonstration of the classification of surrogacy relationship according to IQWiG



14. Kemp R, et al. *BMC Med*. 2017;15(1):1-7.

Methods: Treatment-effect surrogacy

- Patients were allocated to 12 non-overlapping clusters with equal sizes of synthetic treatment and control arms (range: 9-23 patients) using propensity-score matching to balance the differences in baseline age, race, sex, ECOG, and BRAF mutation status
- For each surrogacy relationship, Pearson's correlation coefficient (*r*) between the treatment effects was estimated through a surrogacy equation regressing the log-transformed hazard ratios for SEs (log HR_{SE}) on the log-transformed HRs for OS (log HR_{OS}) across the clusters, weighted by their sample sizes
- For each association, 95% CI for the *r* was estimated using bootstrapping
- Surrogate threshold effect (STE), the minimum treatment effect on the SE to translate into significant OS benefit at a default 95% confidence level,[15] was estimated for PFS-OS and TNTD-OS relationships
- Strength of association between the treatment effects was also assessed using IQWiG[14]

14. Kemp R, et al. *BMC Med*. 2017;15(1):1-7. 15. Burzykowski T, et al. *Pharm Stat*. 2006;5(3):173-186.

Methods: Treatment-effect surrogacy – Validation

- Internally validated the impact of patient allocation to clusters on the results by re-calculating *r* using 10 alternative sets of allocations
- Internally validated the surrogacy equation using leave-one-out cross validation (LOOCV), in which a new model was fitted to the data after omitting one cluster and then used to predict HR_{OS} from HR_{PFS} for that particular cluster
- Externally validated for the PFS-OS surrogacy equation (based on secondary definition of PFS) on an extensive set of RCTs identified in a previous systematic literature review of previously untreated mMel[7]
  - Evidence base included 23 RCTs published between 2000 to 2020
  - Majority of the trials were phase II and III (21 of 23 trials), and multinational (20 of 23) with one trial in the US only
  - Ten trials consisted of at least one treatment arm with immunotherapy or a combination with an immunotherapy agent, whereas five trials consisted of comparisons of immunotherapy agents
  - Seven trials either did not permit treatment crossover or reported crossover-adjusted efficacy estimates

7. Leung L, et al. *Value Health* 2022;25(1):S22.

Results: Patient selection

- A total of 489 patients met the eligibility criteria. All 489 patients were included in the individual-level surrogacy analyses, whereas 428 of them were included in the treatment-effect surrogacy analysis
- Race was majority White, gender was majority male, the median baseline age was 65.6 years, the predominant stage at diagnosis was IV, and the most common ECOG score was zero
- The most commonly administered 1L therapies were ipilimumab, nivolumab, or their combination with a joint share of 52.9%
- Among all eligible patients, no patient initiated 1L treatment before 2011. Most patients (68.1%) initiated 1L therapy after 2015. BRAF mutation status was almost evenly distributed between mutant and wild-types

Variable	Level	Proportion
Gender	Female	36.4%
Race	White	90.2%
BRAF mutation status	Wild-type	53.2%
Cancer stage	IV	58.5%
ECOG	0	53.0%
	1	34.8%
	2+	12.3%
Year of 1L treatment initiation	>2015	68.1%
1L therapy	Dabrafenib + trametinib	10.0%
	Ipilimumab	15.1%
	Ipilimumab + nivolumab	20.4%
	Nivolumab	17.4%
	Pembrolizumab	20.2%
	Other[a]	16.8%

[a] The "Other" category consisted of the following: cobimetinib + nivolumab; dabrafenib + ipilimumab + nivolumab + trametinib; dabrafenib; dabrafenib + pembrolizumab; dabrafenib + trametinib + vemurafenib; dacarbazine + ipilimumab; ipilimumab + vemurafenib; cobimetinib + vemurafenib; methotrexate + pembrolizumab; nivolumab + talimogene laherparepvec; paclitaxel protein-bound; rituximab-PVVR; talimogene laherparepvec; peginterferon alfa-2B; vemurafenib; binimetinib + encorafenib; temozolomide; interferon alfa-2B; trametinib; aldesleukin; anastrozole + palbociclib; carboplatin + paclitaxel; carboplatin + paclitaxel protein-bound; and cisplatin + dacarbazine + vinblastine.

Results: Individual-level surrogacy

- PFS, TNTD and TTD were all moderately associated with OS with their corresponding Spearman's ρ ranging from 0.60–0.67
- The correlation between TTP and OS was weak with a corresponding Spearman's ρ of 0.47
- Considering 2L treatment initiation as an event for the alternative PFS definition had negligible impact on its association with OS
- Compared to other candidate SEs, TNTD exhibited a stronger individual-level correlation with OS. It was followed by PFS, TTD and TTP (in descending order of strength)

Surrogate	ρ (95% CI)	Strength
PFS	0.62 (0.54, 0.72)	Moderate
TNTD	0.67 (0.56, 0.76)	Moderate
TTD	0.60 (0.43, 0.74)	Moderate
TTP	0.47 (0.36, 0.58)	Weak

CI, confidence interval; OS, overall survival; PFS, progression-free survival; TNTD, time-to-next-treatment or death; TTD, time-to-treatment discontinuation; TTP, time-to-progression.

Results: Treatment-effect surrogacy

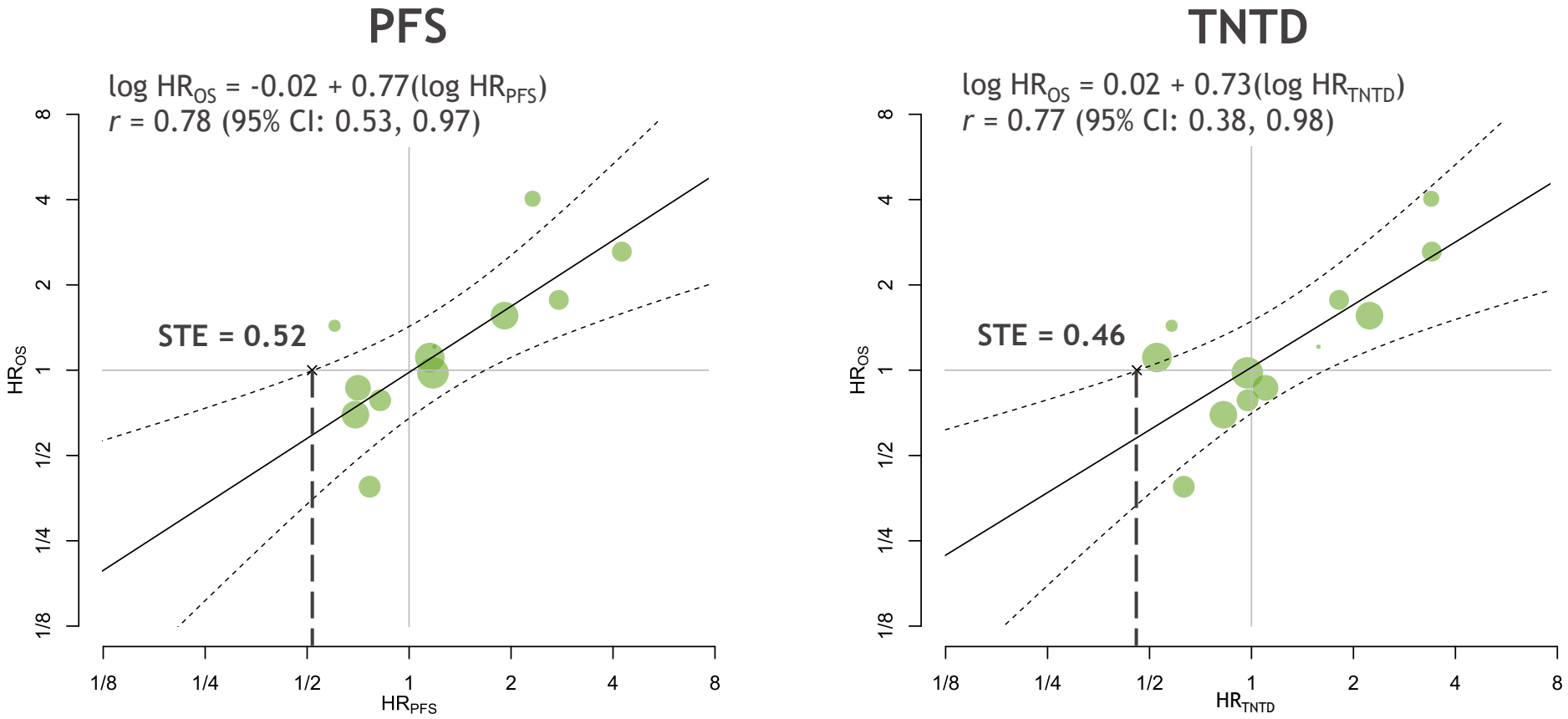
- Correlation between the log-transformed treatment effects (log HR_{SE} and log HR_{OS}) was moderate for all SEs
- Internal validity using LOOCV had accuracy of >90% for TNTD and TTD, and 83.3% for PFS and TTP
- Defining 2L treatment initiation as an event in PFS definition had negligible impact on the strength of its association with OS
- Compared to other candidate SEs, TTD showed stronger treatment-effect correlation with OS. It was followed by PFS, TNTD and TTP (in descending order of strength)

Surrogate	*r* (95% CI)	Surrogacy Equation	LOOCV Accuracy
PFS	0.78 (0.53, 0.97)	$\log HR_{OS} = -0.02 + 0.77(\log HR_{PFS})$	83.3%
TNTD	0.77 (0.38, 0.98)	$\log HR_{OS} = 0.02 + 0.73(\log HR_{TNTD})$	100.0%
TTD	0.87 (0.69, 0.97)	$\log HR_{OS} = 0.06 + 0.63(\log HR_{TTD})$	91.7%
TTP	0.67 (0.36, 0.97)	$\log HR_{OS} = 0.09 + 0.57(\log HR_{TTP})$	83.3%

CI, confidence interval; HR, hazard ratio; LOOCV, leave-one-out cross validation; OS, overall survival; PFS, progression-free survival; TNTD, time-to-next-treatment or death; TTD, time-to-treatment discontinuation; TTP, time-to-progression.

Results: Treatment-effect surrogacy – PFS-OS and TNTD-OS

- As the slopes of the estimated surrogacy equations for PFS and TNTD are similar (0.77 vs. 0.73), unit improvements in the HR_{PFS} or HR_{TNTD} are expected to generate similar magnitude of improvements on HR_{OS}



The weighted linear regression is graphed as a solid straight line with its corresponding 95% prediction interval as dotted curves. Green circles represent synthetically generated clusters from the data. Sizes of the circles are proportional to the corresponding numbers of patients in the clusters. The 95% prediction interval was generated using the median sample size of the RCTs (n = 423) covered in the evidence base of Leung et al. (2022).[7]

CI, confidence interval; HR, hazard ratio; OS, overall survival; PFS, progression-free survival; STE, surrogate threshold effect; TNTD, time-to-next-treatment or death.

7. Leung L, et al. *Value Health* 2022;25(1):S22.

Results: Treatment-effect surrogacy – Internal validity

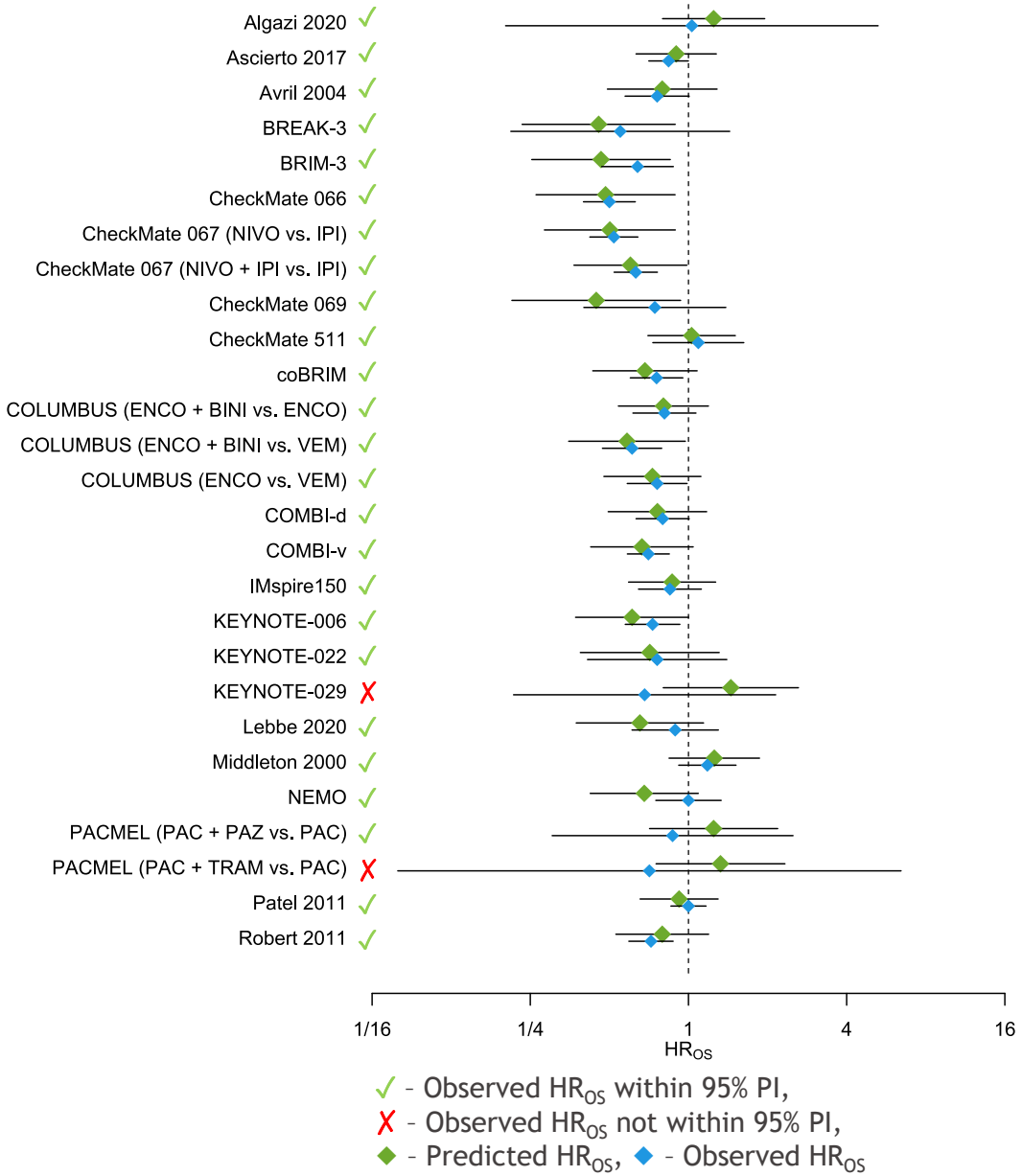
- For each SE, sensitivity analyses were conducted using ten new allocations of patients to clusters
- The 95% CIs of the correlation coefficients estimated for the base case analysis covered all but 4 of the 40 (90%) correlation coefficients estimated in sensitivity analysis (exceptions: one for TTP and three for TTD fell outside of 95% CIs)
- The median of the 10 alternative correlation estimates was slightly (by 0.03–0.05) higher than its base case counterpart for PFS, TNTD, and TTP, but by 0.14 lower for TTD

Surrogate	Base Case Analysis	Sensitivity Analysis
	***r* (95% CI)**	**Median *r* (Range)**
PFS	0.78 (0.53, 0.97)	0.81 (0.49, 0.93)
TNTD	0.77 (0.38, 0.98)	0.82 (0.76, 0.90)
TTD	0.87 (0.69, 0.97)	0.73 (0.40, 0.81)
TTP	0.67 (0.36, 0.97)	0.72 (0.25, 0.88)

CI, confidence interval; PFS, progression-free survival; TNTD, time-to-next-treatment or death; TTD, time-to-treatment discontinuation; TTP, time-to-progression.

Results: Treatment-effect surrogacy – External validity

- The predictive accuracy of the surrogacy equation developed for PFS-OS correlation (using the secondary definition of PFS) was assessed by comparing the observed HR_{OS} in 27 treatment comparisons from 23 RCTs[7] to their corresponding 95% prediction intervals (PIs)
- The observed HR_{OS} was within its 95% PI in 25 (93.6%) comparisons
- In the two comparisons where observed HR_{OS} fell outside of their 95% PIs generated from the model, observed improvements in PFS and OS were discordant (i.e., $HR_{PFS\ secondary\ definition} > 1$ and $HR_{OS} < 1$)
- The predictions on the statistical significances of HR_{OS} matched the results reported from the trials in 20 (74.1%) comparisons between the 95% PI and the 95% CI



7. Leung L, et al. *Value Health* 2022;25(1):S22.

Discussion: Comparison to past research in RWD and RCT settings

- Estimated treatment-effect correlation between PFS (based on secondary definition) and OS from RWD (0.77; 95% CI: 0.53, 0.95) was similar to a previously published estimate from a correlation meta-analysis of aggregate-level RCT data (0.79; 95% CI: 0.59, 0.91)[7]
- Estimated individual-level correlations for PFS, TNTD and TTD, and their corresponding 95% CIs were similar to the estimates obtained in Mohr et al. (2022)[11] on the 664 pembrolizumab-treated advanced melanoma patients from a German registry
- On the individual level, TNTD-OS correlation was consistently stronger than PFS-OS correlation across the current study, Larkin et al. (2022),[10] and Mohr et al. (2022)[11]

Surrogate	ρ (95% CI)			r^2 (95% CI)	
	Mohr et al. (2022)	Larkin et al. (2022)	Current Study	Larkin et al. (2022)	Current Study[a]
PFS	0.52 (0.44, 0.59)	0.72 (0.70, 0.73)	0.62 (0.54, 0.72)	0.71 (0.23, 1.00)	0.61 (0.28, 0.94)
TNTD	0.70 (0.65, 0.75)	0.77 (0.76, 0.78)	0.67 (0.56, 0.76)	0.75 (0.32, 1.00)	0.59 (0.14, 0.96)
TTD	0.62 (0.55, 0.68)	-	0.60 (0.43, 0.74)	-	0.76 (0.48, 0.94)

[a] Calculated from squaring the correlation estimates

CI, confidence interval; PFS, progression-free survival; TNTD, time-to-next-treatment or death; TTD, time-to-treatment discontinuation.

7. Leung L, et al. *Value Health* 2022;25(1):S22. 10. Larkin J, et al. *Ann Oncol* 2022; 33(7):S919-S920. 11. Mohr P, et al. *Cancers* 2022;14(7):1804.

Conclusion

- All individual- and treatment-effect correlations were moderate, except for the endpoint-level correlation for TTP. This behavior can be attributable to the definition of TTP which does not include death as an event
- External validation of the surrogacy equation between the secondary definition of PFS and OS showed high predictive accuracy for RCT data. Estimated strength of the PFS-OS correlation from this RWD was consistent with that from past research in RWD and RCT settings
- Since correlations were labeled as “moderate” per IQWiG guidelines, more work is needed to conclusively validate these surrogates
- Predictions from the derived surrogacy equations may enable earlier assessment of real-world treatment effect on OS without having to collect or procure OS data, which may be costly
- Analysis of RWD on surrogate predictors can complement the existing evidence from clinical trials by increasing its generalizability to more inclusive patient populations, as well as to realistic settings that reflect patients’ and physicians’ treatments in day-to-day practice

Limitations

- The Flatiron Health database may not represent the entire US advanced melanoma population
- The use of RWD may be prone to bias in endpoint measurements due to lack of systematic control over data collection
- We have not demonstrated a causal relationship between each SE and OS, which would require a biological plausibility study
- Caution should be taken when applying the surrogacy models from this work to predict survival outcomes in RCT settings, since (1) correlations were only moderate with wide 95% CIs, and (2) due to differences between the patient populations in real-world and RCT settings
- Due to limited data per treatment basis, in some of the clusters, synthetic treatment or control arms consisted of patients receiving different treatments
- Wide 95% PIs in the external validation were observed as a result of low sample sizes of clusters

Thank you!

Presenter Contact Information
Lisa Leung: lleung@evidinno.com