Comparative Analysis of Nivolumab and Nivolumab Plus Ipilimumab in MSI-H Metastatic Colorectal Cancer, A Bayesian Borrowing and Quantitative Bias Approach

CO135

Dixon, M.[1], Chen, H.[1], Hsu, G.[2], Yajnik, P.[2], Tang, F.[2], Haider, M.[2]

[1]Bristol Myers Squibb, Lawrenceville, NJ, United States; [2]Cytel Inc., Waltham, MA, United States

Background and Objectives

- DNA mismatch repair (MMR) gene mutations which result in MMR deficiency (dMMR) and microsatellite instability (MSI) are linked to (OR occur in) approximately 5% of metastatic colorectal cancer (mCRC) patient cases [1][2].
- MSI-high (MSI-H) colorectal tumors typically exhibit poor differentiation, high lymphocyte infiltration, and mucinous histology [3].
- In 2014, Bristol-Myers Squibb (BMS) launched the CheckMate142 study (NCT02060188), a phase II, multicohort, open-label study assessing nivolumab monotherapy and nivolumab plus ipilimumab in metastatic CRC (mCRC).
- The lack of randomization in CheckMate142 resulted in discrepancies in patient characteristics across cohorts, which may bias treatment effects [4].
- Statistical adjustments for key confounders can help mitigate bias in effect estimates.
- Limited cohort sizes may lead to inconclusive results due to low statistical power and precision.
- Bayesian borrowing was employed to enhance effect estimates by incorporating data from a third cohort receiving nivolumab plus ipilimumab as first-line treatment [5][6].
- The robustness of findings was evaluated using quantitative bias analyses (QBA) to quantify potential biases affecting study validity [7][8][9]:
  - **Uncertain MSI-H status**: to assess the impact of potentially misclassified MSI-H status on treatment outcomes due to errors in testing
  - **Unmeasured confounding**: to assess the strength of an unmeasured confounder linked to both the exposure and outcome required to nullify the observed results
  - **Missing data**: to assess how various missing data assumptions may impact the observed results
  - **Target cohort differences due to Bayesian borrowing**: to address concerns of borrowed patients potentially having inflated survival due to being on different lines of treatment

Methods

- Head-to-head comparative effectiveness study for 2L+ nivolumab monotherapy vs. 2L+ nivolumab + ipilimumab (nivo+ipi) combination therapy
- **Population**: Adult patients with MSI-H/dMMR mCRC enrolled in the CheckMate 142 trial (NCT02060188)
- **Endpoints of interest**: OS and PFS
- **Pre-processing**: Approximately balanced the distributions of potential confounders at baseline using cardinality matching.
- **Bayesian borrowing**: Improved cohort size by borrowing from a cohort of patients who initiated nivolumab + ipilimumab combination as 1L treatment in the CheckMate 142 study to augment the limited sample sizes. The Bayesian analysis employed a power prior, which allows borrowing pre-specified amounts of information from Cohort 3 into the analysis.
- **Time-to-event outcome models**:
  - Parametric (Weibull and piece-wise exponential) proportional hazards models
- **QBA**: Assessed the robustness of the study findings by quantifying the impact of multiple sources of bias
  - Uncertain MSI-H status: Sensitivity samples were created by randomly removing 15% and 25% of patients with only a local test or false positive local tests. The "worst" and "best" samples were identified for Bayesian borrowing.
  - Unmeasured confounding: The sensitivity of results to potential bias from known and unknown unmeasured variables was assessed for significant hazard ratios using the E-value, which was contextualized with study data.
  - Missing data: KRAS/BRAF mutation status was identified as a key confounder, and missing data were analyzed to determine if study conclusions remain valid under all plausible missingness assumptions.
  - Target cohort differences due to Bayesian borrowing: Borrowing 1L patients from Cohort 3 into a 2L+ may introduce bias favoring nivo+ipi. A tipping point analysis was conducted to determine how much OS and PFS in Cohort 3 can be worsened using a multiplicative constant before effect estimates lose statistical significance.

Results

Overall Survival (OS) - Bayesian Borrowing using power priors

- Nivo+ipi patients have significantly better OS compared with nivo patients
- Estimated HRs and CIs are not very sensitive to the amount of borrowing from Cohort 3, with this most likely owing to a combination of factors, chiefly:
  - The relatively large size of Cohort 2 (nivo+ipi, 2L+) compared with Cohort 1 (nivo, 2L+) after pre-processing (119 vs 55)
  - Being left with 35 patients in Cohort 3 (nivo+ipi, 1L) after pre-processing

Table 1. OS estimates from the Weibull model

Power Prior Borrowing Weight[1]	HR
0	0.469 [0.293, 0.757]
0.25	0.467 [0.293, 0.746]
0.5	0.465 [0.294, 0.740]
0.75	0.463 [0.297, 0.736]
1	0.462 [0.295, 0.732]

1. 0 implies no pooling, 1 implies Cohort 3 fully pooled into Cohort 2

Progression Free Survival (PFS) - Bayesian Borrowing using power priors

- The Weibull model was considered less appropriate given the poor fit with the data, regardless of how much borrowing was done (Figure 1)

Figure 1. Full Pooling Weibull Proportional Hazards Model Fit

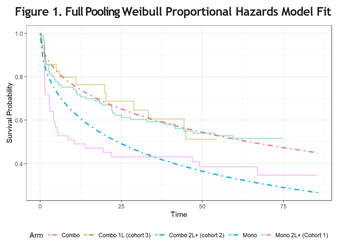


Table 2. PFS estimates from the Weibull model

Power Prior Borrowing Weight[1]	HR
0	0.603 [0.392, 0.941]
0.25	0.604 [0.396, 0.940]
0.5	0.605 [0.399, 0.938]
0.75	0.606 [0.399, 0.935]
1	0.603 [0.402, 0.927]

1. 0 implies no pooling, 1 implies Cohort 3 fully pooled into Cohort 2

Figure 2. Full Pooling PWE Proportional Hazards Model Fit

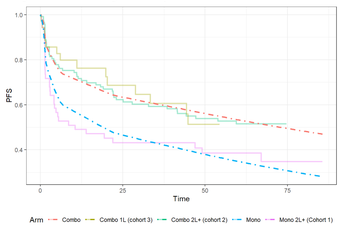


Table 3. PFS estimates from the PWE model

Power Prior Borrowing Weight[1]	HR
0	0.597 [0.389, 0.933]
0.25	0.596 [0.391, 0.928]
0.5	0.596 [0.391, 0.922]
0.75	0.596 [0.395, 0.919]
1	0.596 [0.395, 0.915]

1. 0 implies no pooling, 1 implies Cohort 3 fully pooled into Cohort 2

- Using a piece-wise exponential (PWE) model, better fit (Figure 2) than the Weibull model was achieved. Nivo+ipi patients had significantly better PFS compared with nivo patients
- Estimated HRs were not very sensitive to the amount of borrowing, with this most likely owing to the same combination of factors as for the OS analyses

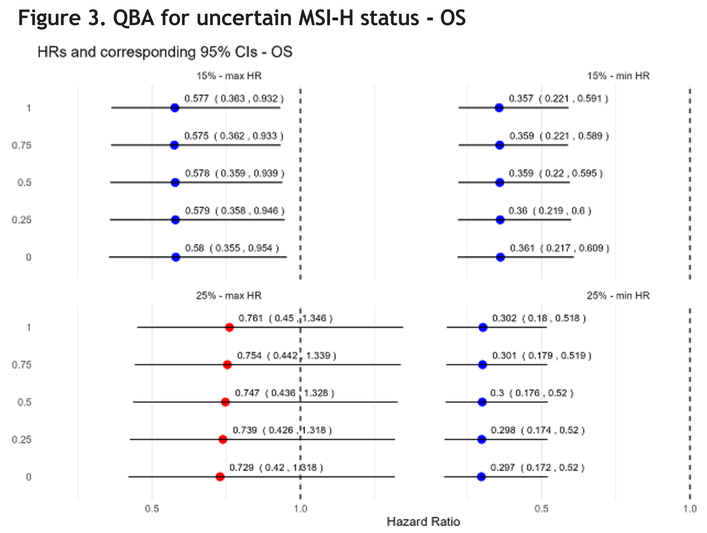
Quantitative bias analysis results

- Study conclusions are largely either very robust or generally robust for OS and PFS against all the sources of bias assessed:

1. Uncertain MSI-H status

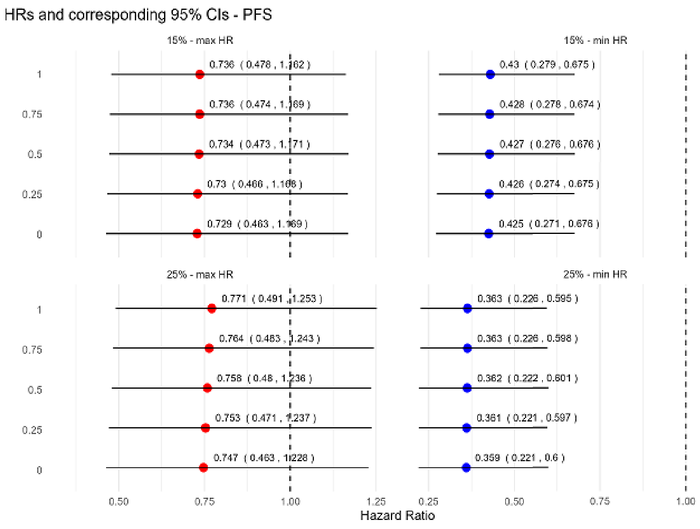
- QBA for patients with only a local test or a false positive local test – OS **(Figure 3)**
  - After randomly removing 15% (top panels) or 25% (bottom panels) of patients:
  - In the worst-case scenario (left panels), i.e. subset with the worst possible HR estimates:
    - **15% scenario:** estimates are fully robust under all borrowing scenarios with respect to the CIs and point estimates
    - **25% scenario:** estimates are robust with respect to point estimates but are not robust with respect to CIs under all borrowing scenarios
    - Note that the lack of robustness for CIs likely does not largely impact the robustness of the study conclusions in general as the CIs are very wide, suggesting that the perceived lack of robustness is owing to poor balance between cohorts (leading to small cohort sizes and thus power)
  - In the best-case scenario (right panels), i.e. subset with the best possible HR estimates, the estimates are **fully robust** under all borrowing scenarios with respect to the CIs and point estimates
  - **For OS, the study results are quite robust against patients with only a local test or a false positive local test under all borrowing scenarios**

Figure 3. QBA for uncertain MSI-H status - OS



- QBA for patients with only a local test or a false positive local test – PFS **(Figure 4)**
  - After randomly removing 15% (top panels) or 25% (bottom panels) of patients:
  - In the worst-case scenario (left panels), i.e. subset with the worst possible HR estimates, the effect estimates are
    - Robust with respect to the point estimates
    - Not robust under any borrowing scenarios with respect to the CIs,
    - Note that as with OS, the lack of robustness for CIs is to some extent owing to poor balance between cohorts (leading to small cohort sizes and thus power)
  - In the best-case scenario (right panels), i.e. subset with the best possible HR estimates, the estimates are fully robust under all borrowing scenarios with respect to the CIs and point estimates
  - **For PFS, the study results may not be robust against patients with only a local test or a false positive local test under all borrowing scenarios**
    - **QBA suggest that a relatively large degree of borrowing can improve the robustness of study results**

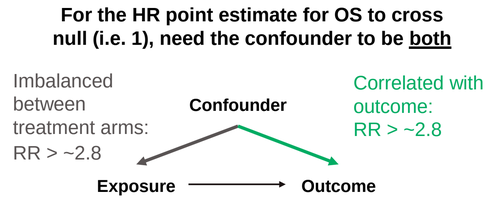
Figure 4. QBA for uncertain MSI-H status - PFS



2. Unmeasured confounding (Figure 5)

- E-value: the minimum strength of association that an unmeasured confounder would need to have with both the treatment and the outcome to fully explain away a specific treatment-outcome association, conditional on the measured covariates
- Based on the associations of all measured confounders with treatment/outcomes, one fairly strong associations between unmeasured confounders and treatment/outcomes are needed to nullify observed point estimates [OS (RR>~2.8); PFS (RR>~2.2)]
- Generally strong associations are also required to nullify the observed upper CIs [OS (RR>~1.8); PFS (RR>~1.3)]
- Thus, the study conclusions
  - **For OS are very robust against unmeasured confounding**
  - **For PFS are generally robust against unmeasured confounding**

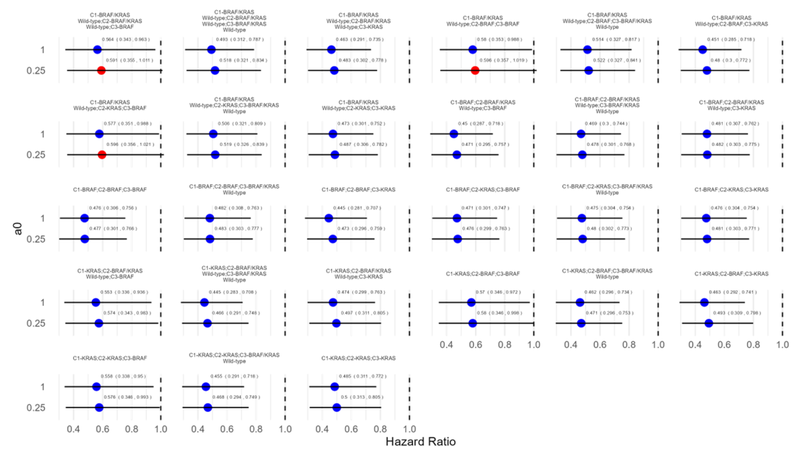
Figure 5. QBA for unmeasured confounding



3. QBA for missing data - BRAF/KRAS mutation

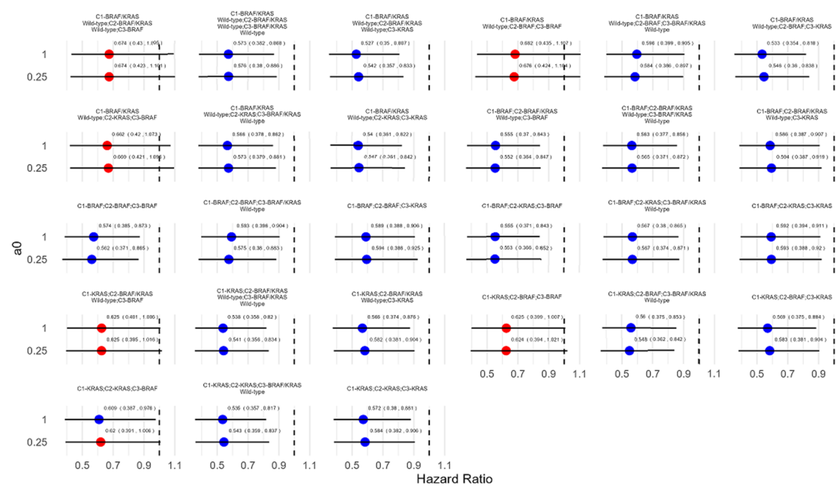
- **OS (Figure 6)**
  - **The study results for OS are very robust against missing BRAF/KRAF mutation status, particularly as the amount of borrowing increases.**
  - All HR point estimates of OS are < 1 across all scenarios
  - All upper CIs are < 1 across all scenarios under full borrowing
  - The upper CI very slightly crosses 1 for only 3 out of 27 scenarios under partial borrowing

Figure 6. HR and corresponding 95% CIs with imputed BRAF/KRAS mutation status - OS



- **PFS (Figure 7)**
  - **The study results for PFS are generally robust against missing BRAF/KRAF mutation status.**
  - All HR point estimates of OS are < 1 across all scenarios
    There are 5 scenarios where the upper CI crosses 1 for both borrowing cases
  - Unlike in the QBA for uncertain MSI-H status, PFS robustness against mutation status does not seem as affected by the amount of borrowing

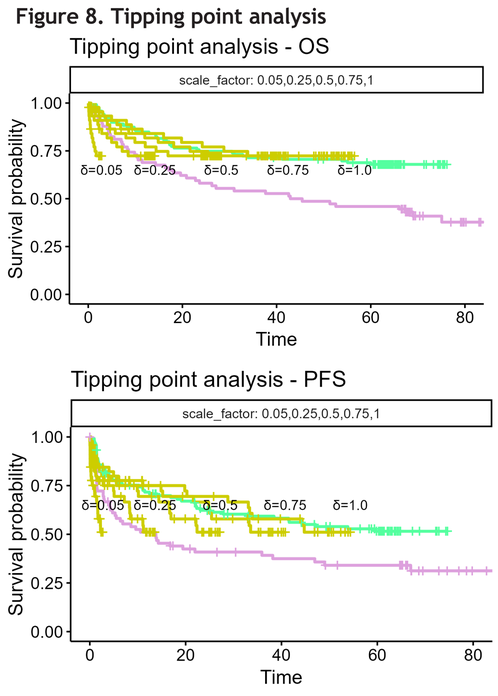
Figure 7. HR and corresponding 95% CIs with imputed BRAF/KRAS mutation status - PFS



4. QBA for target cohort differences due to Bayesian borrowing (Figure 8)

- Figure 8 displays incremental KM curves with varying scale factors (0 to 1) for OS and PFS, showing the impact of scaled-down survival assumptions on survival probabilities
- With respect to HR point estimates, no tipping points where conclusions were reversed were identified for OS and PFS
- With respect to HR upper CIs, the tipping point where conclusions were reversed was found for only PFS scenarios that were most likely implausible, i.e. when patients had progression times reduced by over 90%

Figure 8. Tipping point analysis



Scale factor = 1: no adjustment made on OS of Cohort 3
Scale factor = 0.5: artificially shorten OS of Cohort 3 by half

Scale factor = 1: no adjustment made on PFS of Cohort 3
Scale factor = 0.5: artificially shorten PFS of Cohort 3 by half

Conclusions

- Based on this analysis of Checkmate 142, patients using nivo+ipi have significantly better OS and PFS compared with patients using nivo.
- QBA shows that study conclusions are largely either very robust or generally robust for OS and PFS against all the sources of bias assessed: uncertainty in MSI-H status, unmeasured confounding, missing data, target cohort differences due to Bayesian borrowing.
- Additional patients would improve the QBA analysis, as it would enhance the statistical power and allow for more interpretable confidence intervals, given that a limitation of the study were the small sample sizes and number of events, thereby allowing us to better assess the robustness of our findings.
- Additional patients would also support the core analyses and QBA in particular for PFS, which has relatively higher uncertainty in main effect estimates, and is relatively less robust compared with OS.

References

1. Yao, Y. C., Jin, Y., Lei, X. F., Wang, Z. X., Zhang, D. S., Wang, F. H., ... & Wang, F. (2022). Impact of mismatch repair or microsatellite status on the prognosis and efficacy to chemotherapy in metastatic colorectal cancer patients: A bi-institutional, propensity score-matched study. Journal of Cancer, 13(9), 2912.
2. Boland CR GA. Microsatellite Instability in Colorectal Cancer. Gastroenterology. 2010;138(6):2073-87.
3. Raut CP PT, Rodriguez-Bigas MA. . Clinicopathologic features in colorectal cancer patients with microsatellite instability. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis. 2004;568(2):275-82.
4. Overman MJ, Lenz H-J, Andre T, Aglietta M, Wong MK, Luppi G, et al. Nivolumab (NIVO)±ipilimumab (IPI) in patients (pts) with microsatellite instability-high/mismatch repair-deficient (MSI-H/dMMR) metastatic colorectal cancer (mCRC): Five-year follow-up from CheckMate 142. American Society of Clinical Oncology; 2022.
5. Viele K, Berry S, Neuenschwander B, Amzal B, Chen F, Enas N, et al. Use of historical control data for assessing treatment effects in clinical trials. Pharm Stat. 2014;13(1):41-54.
6. Muehlemann N, Zhou T, Mukherjee R, Hossain MI, Roychoudhury S, Russek-Cohen E. A Tutorial on Modern Bayesian Methods in Clinical Trials. Ther Innov Regul Sci. 2023;57(3):402-16.
7. Thorlund K, Duffield S, Popat S, Ramagopalan S, Gupta A, Hsu G, et al. Quantitative bias analysis for external control arms using real-world data in clinical trials: a primer for clinical researchers. J Comp Eff Res. 2024:e230147.
8. Lash TL, Fox MP, Fink AK. Applying Quantitative Bias Analysis to Epidemiologic Data: Springer New York; 2011.
9. Wilkinson S, Gupta A, Scheuer N, Mackay E, Arora P, Thorlund K, et al. Assessment of alectinib vs ceritinib in ALK-positive non-small cell lung cancer in phase 2 trials and in real-world data. JAMA Network Open. 2021;4(10):e2126306-e.

Acknowledgments

- Please refer to Scientific Publication Standards (BMS-OCD-1cS1) for acknowledgment requirements

Presented at ISPOR Europe 2024; November 17-20, 2024; Barcelona, Spain

Email: <grace.hsu@cytel.com>