REGORAFENIB IN METASTATIC COLORECTAL CANCER: COST-EFFECTIVENESS ANALYSIS BASED ON PROPENSITY SCORE WEIGHTED COHORT OF CZECH REGISTRY

Tomas Mlcoch,¹ Tereza Hrnciarova¹, Jakub Zadak², Šárka Veselá², Marisca Marian³, Tomas Dolezal¹

¹VALUE OUTCOMES, Prague, Czech Republic;  
²Bayer s.r.o., Prague, Czech Republic;  
³Bayer Pharmaceuticals, Basel, Switzerland.
Introduction

• Regorafenib (Stivarga) is indicated, among others, for treatment of metastatic colorectal cancer (mCRC) after failure of previous therapy for metastatic disease

• If highly innovative by pre-defined criteria, a drug can receive 2+1 (3)* years of temporary reimbursement in the Czech Republic
  
  • Cost-effectiveness analysis is then only informative and not mandatory for a decision. After 2 (3) years, a company has to prove CE, below willingness-to-pay (WTP) threshold otherwise lose reimbursement
  
  • There is a mandatory data collection within temporary reimburs.
  
  • Details of this scheme are well-described by Ornstova et al. (2018)

• Regorafenib was deemed as highly innovative and received temporary reimbursement from **July 2015 to June 2018** (3 years)

• WTP threshold is equal to approx. €47,000/QALY in the Czech Republic (1,2 million CZK)

---

*Initial 2 years can be prolonged by 1 year if: a therapy is still highly innovative (no new therapy) and a company applies for this scheme.
Registry data vs. clinical trial data

• In line with randomized clinical trial (RCT) of regorafenib (CORRECT), there is a trend of higher efficacy in patients with less previous treatment (tx.) lines (Grothey et al.)
  • In CORRECT trial, approx. 50% of patients had ≥4 tx. lines with only 26% having only 1-2 prior lines on or after metastatic disease

<table>
<thead>
<tr>
<th>Previous treatment lines</th>
<th>n</th>
<th>HR OS (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>≤3</td>
<td>301</td>
<td>0.71 (0.52–0.97)</td>
</tr>
<tr>
<td>&gt;3</td>
<td>459</td>
<td>0.80 (0.62–1.04)</td>
</tr>
</tbody>
</table>

• Conversely, in Czech Registry, 69 % of patients received only 1-2 prior lines

• Registry data showed improved overall survival (OS) and progression-free survival (PFS) compared to RCT
  • but was it due to regorafenib efficacy or simply better patient prognosis?
Global pharmacoeconomic model and its evidence

- Global partitioned survival model with 3-health states was used (without progression, progression and death)
  - Utility values: 0.71 (pre-progression); 0.59 (post-progression)
  - Costs of usual care (negligible; ≈€50-80/cycle)*

- The model was based on randomized clinical trial (RCT) CORRECT
  - It followed long-term Kaplan-Meier curves directly (no need to extrapolate)
  - Longer follow-up from RCT was provided (2.5 years longer cut-off date compared to initial study by Grothey et al.)

- However, the model should mimic target population and real-world evidence (RWE) which was collected over 3y of temporary reimbursement

* Exchange rate 25.364 CZK per 1 EUR (as of April 2018).
Global pharmacoeconomic model and its evidence II

• So where to get next: how to reliably incorporate local evidence into the global model?
  • Without controlled (placebo) arm? X
  • With completely different characteristics between placebo RCT data and registry intervention (regorafenib) data? X
Propensity score methods

• Propensity score (PS) weighting balances patient characteristics so as to mitigate effect of chosen variables (Phillipo et al. 2015)

• Then, all outcomes (OS/PFS) are weighted accordingly

• The cohort can be weighted so as to „match“:
  • Initial cohort (in our case RCT)
  • New cohort (in our case registry)
  • Overlap cohort which is artifically estimated so as to achieve the highest overlap between two populations if there are significant differences between two populations (Li et al. 2016)

• These methods can be employed only if individual patient data (IPD) are available for both cohorts
Propensity score weighting

- Regorafenib in registry lacked „control arm“ for efficacy estimation, so we weighted:
  - Registry data - regorafenib arm
  - RCT data - placebo arm

- Best performance was for **PS weighting with overlap weights**
  - PS estimated via logistic regression with binary variable for registry/RCT

- We included parameters that:
  - Could impact treatment effect (e.g. 1-2 prior lines)
  - Have effect on OS/PFS (e.g. age, ECOG)
  - Interactions of above mentioned parameters (if relevant)
  - Have different variability (e.g. age)

- The final model was chosen based on all available variables and 2nd order interactions with hierarchy (marginal effects were preserved)
  - Highly non-significant variables were excluded (p-value from LR test > 0.2)
  - Akaike information criteria (AIC) assessed, also extreme effect of few variables was avoided
  - Final weights distribution and effective sample size (ESS) were taken into account
**Registry and trial data differences**

- There were significant differences between RCT and registry data, some notable differences in red; overlap population from PS in blue
  - In registry, patients had i) fewer prior lines, ii) worse ECOG and iii) lower KRAS mut.

<table>
<thead>
<tr>
<th></th>
<th>CORRECT trial (regorafenib)</th>
<th>CORRECT trial (placebo)</th>
<th>Registry data (regorafenib)</th>
<th>Overlap population</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of patients</td>
<td>505</td>
<td>251</td>
<td>429</td>
<td>-</td>
</tr>
<tr>
<td>Patients included in PS (ESS)</td>
<td>-</td>
<td>-</td>
<td>247* (105)</td>
<td>-</td>
</tr>
<tr>
<td>Age in years (mean)</td>
<td>60.7</td>
<td>60.0</td>
<td>63.8</td>
<td>61.5</td>
</tr>
<tr>
<td>Time from metastases to tx. initiation (months)</td>
<td>35.0</td>
<td>34.6</td>
<td>31.8</td>
<td>31.5</td>
</tr>
<tr>
<td>Number of prior lines for metastatic disease 1-2 (%)</td>
<td>16.2 %</td>
<td>15.1 %</td>
<td>68.8 %</td>
<td>44.2 %</td>
</tr>
<tr>
<td>Number of prior lines for met. disease 3 (%)</td>
<td>24.0 %</td>
<td>23.5 %</td>
<td>22.6 %</td>
<td>31.7 %</td>
</tr>
<tr>
<td>Number of prior lines for met. disease 4 (%)</td>
<td>25.1 %</td>
<td>24.7 %</td>
<td>6.8 %</td>
<td>17.6 %</td>
</tr>
<tr>
<td>Number of prior lines for met. disease &gt;4 (%)</td>
<td>34.7 %</td>
<td>36.7 %</td>
<td>1.9 %</td>
<td>6.5 %</td>
</tr>
<tr>
<td>Male/Female (% male)</td>
<td>61.6 %</td>
<td>59.4 %</td>
<td>62.9 %</td>
<td>61.9 %</td>
</tr>
<tr>
<td>ECOG 0/1 (% with ECOG=1)</td>
<td>47.5 %</td>
<td>43.0 %</td>
<td>64.6 %</td>
<td>57.6 %</td>
</tr>
<tr>
<td>Localization of tumor colon/rectum (% rectum)</td>
<td>35.9 %</td>
<td>33.1 %</td>
<td>38.9 %</td>
<td>27.2 %</td>
</tr>
<tr>
<td>KRAS mutation (% yes)</td>
<td>54.1 %</td>
<td>61.8 %</td>
<td>46.4 %</td>
<td>61.5 %</td>
</tr>
</tbody>
</table>

* In registry, 166 patients were excluded due to missing time from metastases (important outcome); others were excluded due to missing KRAS mutation status.
Propensity score results: logistic regression

• Interpretation of regression results:
  • Estimate KRAS (yes) = -1.652; \( \exp(-1.652) = 0.19 \)
  • A patient has 0.19-times (lower) probability that it will be chosen to registry/treatment arm (than to RCT) compared to a patient without KRAS mutation
  • In **green** higher probability, in **red** lower probability of being chosen to registry

<table>
<thead>
<tr>
<th>Variable*</th>
<th>Estimate</th>
<th>Std. error</th>
<th>Welch t-test p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept (constant term)</td>
<td>-0.831</td>
<td>1.091</td>
<td>0.446</td>
</tr>
<tr>
<td>Age</td>
<td>0.024</td>
<td>0.015</td>
<td>0.093</td>
</tr>
<tr>
<td>ECOG=1</td>
<td>1.762</td>
<td>0.446</td>
<td><strong>&lt;0.001</strong></td>
</tr>
<tr>
<td>Number of lines &gt;4</td>
<td>-4.839</td>
<td>1.563</td>
<td><strong>0.002</strong></td>
</tr>
<tr>
<td>Number of lines =3</td>
<td>-3.302</td>
<td>0.747</td>
<td><strong>&lt;0.001</strong></td>
</tr>
<tr>
<td>Number of lines =4</td>
<td>-5.101</td>
<td>0.966</td>
<td><strong>&lt;0.001</strong></td>
</tr>
<tr>
<td>Localization (rectum)</td>
<td>-0.0003</td>
<td>0.485</td>
<td>0.999</td>
</tr>
<tr>
<td>KRAS mutation (yes)</td>
<td>-1.652</td>
<td>0.323</td>
<td><strong>&lt;0.001</strong></td>
</tr>
<tr>
<td>Male</td>
<td>0.904</td>
<td>0.677</td>
<td>0.182</td>
</tr>
<tr>
<td>Number of lines &gt;4 + time from metastases</td>
<td>0.052</td>
<td>0.020</td>
<td><strong>0.008</strong></td>
</tr>
<tr>
<td>Number of lines 3 + time from metastases</td>
<td>-0.041</td>
<td>0.032</td>
<td>0.198</td>
</tr>
<tr>
<td>Number of lines 4 + time from metastases</td>
<td>0.035</td>
<td>0.024</td>
<td>0.150</td>
</tr>
<tr>
<td>ECOG=1 + male</td>
<td>0.018</td>
<td>0.024</td>
<td>0.458</td>
</tr>
<tr>
<td>Time from metastases + male</td>
<td>-1.177</td>
<td>0.570</td>
<td><strong>0.039</strong></td>
</tr>
<tr>
<td>Localization (rectum) + male</td>
<td>-0.027</td>
<td>0.016</td>
<td>0.097</td>
</tr>
</tbody>
</table>

* Baseline model: ECOG=0, female, no. prior lines 1-2, without KRAS mutation, localization (colon). In **green** higher probability; in **red** lower probability
Propensity score: overlap weights and regression dynamics

- Distribution of weights was similar in both groups and without many extreme values, it provided good ESS (105 and 125 patients)

\[ w_i = \begin{cases} 
\pi_i, & \text{if a patient is from clinical trial (placebo)} \\
1 - \pi_i, & \text{if a patient is from registry (Stivarga)} 
\end{cases} \]

- Dynamics behind propensity score results
  - Registry cohort had significantly fewer previous tx. lines (OS\(\uparrow\)), but
  - There was a higher proportion of ECOG=1 (worse) which had a negative effect (\(\downarrow\) OS/PFS) (see Appendix slides – OS based on ECOG)
Propensity score: overall survival

Unadjusted OS curves

- Placebo (CORRECT) - unadjusted
- Regorafenib (Registry) - unadjusted
Propensity score: overall survival

- HR OS (PS): 0.53 (95% CI: 0.34-0.83)
  - There was a slight shift in both arms (bigger leftward in regorafenib arm)
Propensity score: progression-free survival

Unadjusted PFS curves
Propensity score: progression-free survival

- HR PFS (PS): 0.32 (95% CI: 0.21-0.49)
  - There was a leftward shift in regorafenib arm (placebo remained almost the same)
Updated cost-effectiveness results

- Over time, there was also a huge drop in price based on external price referencing (≈50%; from €5,100 to €2,450/pack)
- Based on PS weighting, the ICER was €43,122 which is below WTP threshold
  - This is due to by more than a 2-times increase in incremental QALYs

<table>
<thead>
<tr>
<th>Cost-effectiveness results based on RCT data (whole population)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter</td>
</tr>
<tr>
<td>Total costs</td>
</tr>
<tr>
<td>Total QALYs</td>
</tr>
<tr>
<td>ICER (€/QALY)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cost-effectiveness results based on propensity score (registry/RCT data)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total costs</td>
</tr>
<tr>
<td>Total QALYs</td>
</tr>
<tr>
<td>ICER (€/QALY)</td>
</tr>
</tbody>
</table>


Conclusions

• Regorafenib showed that it is most likely cost-effective in mCRC patients after less prior lines in the Czech Republic:
  • It is likely that there would be better efficacy if the RCT was designed in this way.

• In registry, regorafenib had markedly higher efficacy than in RCT, which might be caused by:
  • True efficacy (if RCT would had been conducted...)
  • No randomization, and potential confounding and unobservable/unquantifiable effect due to treating patients with better prognosis („physicians know“)
    • PS weighting does not substitute randomization

• PS provides an „evidence bridge“ in situations when:
  • We lack control (placebo) arm
  • There are different patient characteristics between studied populations
Conclusions II

• In the end, State Institute for Drug Control (SUKL) did not take PS weighting into account and insisted on RCT comparison
  • Regorafenib was deemed similarly efficacious to trifluridine/tipiracil and there was an agreement on price parity (both therapies had confidential discounts)
    • Based on NMA Abrahao et al. (2018)

• Takeaway message – although there are some primary endpoints, it is sometimes worth (especially in large diagnosis such as mCRC) to recruiting more patients so as to have better stratification of sub-populations
  • Also, more severe/pre-treated patients ≠ better relative efficacy
Disclosure

• This study was sponsored by Bayer Czech Republic
References


• Fan Li, Kari Lick Morgan, Alan M. Zaslavsky. Balancing Covariates via Propensity Score Weighting
APPENDIX
Progression-free survival from CORRECT trial (longer follow-up)

- Kaplan-Meier curves with 95% confidence intervals (time in months)
  - HR PFS: 0.46 (95% CI: 0.39-0.55)
Survival probability based on ECOG

- Kaplan-Meier curves from CORRECT trial (placebo arm) based on ECOG with 95% confidence intervals (time in months)
Overall survival from CORRECT trial (longer follow-up)

- Kaplan-Meier curves with 95% confidence intervals (time in months)
  - HR OS: 0.79 (95% CI: 0.67-0.92)
Illustration of overlapping population

- Based on 2 parameters: number of prior lines and time from metastases in months
  - Color saturation represents given weight – higher the weight, more satured color
  - To number of prior lines, random number from interval (-0.4; 0.4) was added for better readability of this graph