Indirect treatment comparison of bevacizumab + interferon-α-2a vs tyrosine kinase inhibitors in first-line metastatic renal cell carcinoma therapy

Gerald HJ Mickisch1
Björn Schwander2
Bernard Escudier3
Joaquim Bellmunt4
José P Maroto5
Camillo Porta6
Stefan Walzer7
Uwe Siebert8,9

1Department of Urology, Center of Operative Urology Bremen, Bremen, Germany; 2Department of Outcomes Research, AiM GmbH Assessment-in-Medicine, Lörrach, Germany; 3Immunotherapy Unit, Institut Gustave Roussy, Villejuif, France; 4Department of Medical Oncology, University Hospital del Mar UPF, Barcelona, Spain; 5Department of Medical Oncology, Hospital de la Santa Creu i Sant Pau, Barcelona, Spain; 6Department of Medical Oncology, IRCCS San Matteo University Hospital Foundation, Pavia, Italy; 7Global Health Economics, F Hoffmann-La Roche Pharmaceuticals AG, Basel, Switzerland; 8Department of Public Health, Medical Decision Making and Health Technology Assessment, UMIT - University for Health Sciences, Medical Informatics and Technology, Hall I.T., Austria; 9Department of Health Policy and Management, Harvard School of Public Health, Boston, MA, USA

Correspondence: Gerald HJ Mickisch
Center of Operative Urology Bremen,
Robert-Koch-Str 34a, 28277 Bremen, Germany
Tel +49 421 8700300
Fax +49 421 8700309
Email gerald.mickisch@coub.de

Background: The vascular endothelial growth factor inhibitor bevacizumab (BEV) given in combination with interferon-α-2a (IFN), and the tyrosine kinase inhibitors (TKIs) sunitinib (SUN) and pazopanib (PAZ), have all shown significant increase in progression-free survival (PFS) in first-line metastatic renal-cell carcinoma (mRCC) therapy. These targeted therapies are currently competing to be primary choice; hence, in the absence of direct head-to-head comparison, there is a need for valid indirect comparison assessment.

Methods: Standard indirect comparison methods were applied to independent review PFS data of the pivotal Phase III trials, to determine indirect treatment comparison hazard-ratios (HR) with 95% confidence intervals (95% CI). As BEV+IFN and SUN have been compared to IFN, indirect comparison was enabled by the common IFN comparator arms. As PAZ was compared to placebo (PLA), a connector trial (IFN vs PLA) was required for the indirect comparison to BEV+IFN. Sensitivity analyses taking into account real-life influence of patient compliance on clinical outcomes were performed.

Results: The indirect efficacy comparison resulted in a statistically nonsignificant PFS difference of BEV+IFN vs SUN (HR: 1.06; 95% CI: 0.78–1.45; \( P = 0.73 \)) and of BEV+IFN vs PAZ (range based on different connector trials; HR: 0.74–1.03; \( P = 0.34–0.92 \)). Simulating real-life patient compliance and its effectiveness impact showed an increased tendency towards BEV+IFN without reaching statistical significance.

Conclusions: There is no statistically significant PFS difference between BEV+IFN and TKIs in first-line mRCC. These findings imply that additional treatment decision criteria such as tolerability and therapy sequencing need to be considered to guide treatment decisions.

Keywords: indirect treatment comparison, progression-free survival, renal cell carcinoma, bevacizumab, sunitinib, pazopanib

Introduction

Metastatic renal-cell carcinoma (mRCC) has always been one of the most drug-resistant malignancies1 and the 5-year survival rates remain low at only around 10% and had not improved by 2008.2,3

Over the past two decades, immunomodulating drugs such as interferon-α-2a (IFN) have been the standard first-line mRCC treatment,4 and have been considered the standard comparator in clinical trials.5 Recent advances in understanding the molecular biology of kidney cancer have resulted in the development of drugs that target known molecular pathways which are believed to be important in this disease, such as vascular endothelial growth factors and their receptors.
The vascular endothelial growth factor inhibitor bevacizumab (BEV) given in combination with IFN, and the tyrosine kinase inhibitors (TKIs) sunitinib (SUN) and pazopanib (PAZ), have all shown significant increase in progression-free survival (PFS) in first-line mRCC therapy. These targeted therapies are currently competing to be the primary choice for the first-line therapy of mRCC patients presenting a good or intermediate prognosis. Hence, in the absence of direct head-to-head comparison, there is a need for valid indirect comparison assessment.

**Material and methods**

**Pivotal trial outcomes**

The published Phase III pivotal trial PFS outcomes have been selected as the basis of the indirect treatment comparison (ITC), as these present the highest quality data based on independent central review assessment. Within these pivotal trials BEV+IFN\(^6\) and SUN\(^7\) have each shown a significant increase in PFS vs IFN in first-line mRCC therapy, whereas PAZ has shown a significant PFS increase compared to placebo (PLA),\(^8\) as shown in Figure 1.

The PFS hazard ratios (HRs) were selected as the preferred outcome for the ITC, as this effect measure accounts for censoring and incorporates time to event information.\(^9\)

The independent review PFS HR of BEV+IFN vs IFN is 0.57 (95% confidence intervals [95% CI]: 0.45–0.72; \(P < 0.0001\)),\(^6\) the PFS HR of SUN vs IFN is 0.54 (95% CI: 0.44–0.66; \(P < 0.00001\))\(^7\) and the PFS HR of PAZ vs PLA is 0.40 (95% CI: 0.27–0.60; \(P < 0.001\)),\(^8\) respectively.

The BEV+IFN study named AVOREN and the SUN trial focused on treatment-naïve mRCC patients (first-line population), whereas the PAZ study included both treatment-naïve and pretreated mRCC patients. Hence for the ITC the pazopanib results of treatment-naïve patients have been applied, based on prespecified subgroup analysis.

As shown in Table 1 study designs, patient characteristics, enrollment criteria, and study measurements are comparable, but not identical, between the AVOREN trial, the SUN trial, and the PAZ study.

AVOREN and the PAZ trial were double-blinded placebo-controlled randomized trials, whereas the SUN study was a randomized open-label study. Furthermore, within the AVOREN trial 100% of patients were nephrectomized (inclusion criteria) whereas in the SUN and the PAZ trial 88%–91% of patients had a previous nephrectomy. Another difference is that the SUN and the PAZ trials included more patients with a favorable prognosis (MSKCC risk score 0: 34%–39%) compared to the AVOREN study (27%–29%). Although both factors are regarded as predictive for the PFS outcome, the between-study differences are small, hence performing an indirect treatment comparison (ITC), without applying adjustments for patient characteristics variations, was regarded as an appropriate approach.

**Indirect treatment comparison approach**

The indirect treatment comparison of PFS outcomes of BEV+IFN vs SUN and vs PAZ uses the most widely applied indirect comparison approach by Bucher et al.\(^12\)

---

**Figure 1** Pivotal Phase III progression-free survival outcomes in first-line mRCC therapy.

**Note:** PAZ study results refer to the first-line sub-population.

**Abbreviations:** AVOREN, AVastin fOr RENal cell cancer; BEV, bevacizumab; CI, confidence intervals; IFN, interferon-\(\alpha\)-2a; HR, hazard ratio; PAZ, pazopanib; PFS, progression-free survival; PLA, placebo; SUN, sunitinib.
The Canadian Agency for Drugs and Technologies in Health\textsuperscript{13} and others\textsuperscript{14,15} have recently identified this method as the most suitable approach for performing indirect treatment comparisons of randomized controlled trials.

As BEV+IFN and SUN have been compared to IFN, indirect comparison was enabled by the common IFN control arms, whereas for comparing BEV+IFN vs PAZ a connector trial (IFN vs PLA) is required, as shown in Figure 2.

For the identification of suitable connector trials a systematic literature search was performed using the following literature databases: PubMed, Embase, and Cochrane Central Register of Controlled Trials. As a result only three IFN studies have been identified that provided a suitable PFS HR compared to the Phase III trial outcomes (median PFS of IFN $\approx$ 5 months and median PFS for PLA $\approx$ 3 months) in treatment-naïve mRCC patients. Although none of these compared IFN vs PLA, the selected studies compared IFN regimens either to placebo-like therapy (MRCRCC trial\textsuperscript{16}) or to other IFN regimens that had a placebo-like PFS outcome (Aass et al\textsuperscript{17} and Mickisch et al\textsuperscript{18}). In the absence of a valid IFN vs PLA connector trial all of these studies have been used to perform the ITC of BEV+IFN vs PAZ. Furthermore a PFS HR was estimated (‘proxy comparison’) based on the median PFS time of IFN (5.4 months\textsuperscript{6}) and of placebo (2.8 months\textsuperscript{8}) by assuming constant hazards (HR IFN vs PLA $= 2.8 \text{ m}/5.4 \text{ m} = 0.52$). The selected connector trials and the PFS HRs applied for IFN vs PLA are shown in Table 2.

The indirect comparisons of BEV+IFN vs SUN and BEV+IFN vs PAZ were performed for two key scenarios:
1. Indirect efficacy comparison: comparison of the Phase III results as published.
2. Indirect effectiveness assessment based on simulating the impact of patient compliance.

Table 1 Comparison of the main study design, patient characteristics, enrolment criteria, and study measurements of the underlying pivotal trials

<table>
<thead>
<tr>
<th>Parameter</th>
<th>AVOREN study$^{6,10}$</th>
<th>SUN study$^{7,11}$</th>
<th>PAZ study$^{8}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>BEV+IFN</td>
<td>n = 327</td>
<td>Open-label RCT</td>
<td>PAZ (n = 290)</td>
</tr>
<tr>
<td>IFN</td>
<td>n = 322</td>
<td></td>
<td>PLA (n = 145)</td>
</tr>
<tr>
<td>Study design</td>
<td>Double-blinded placebo controlled RCT</td>
<td></td>
<td>Double-blinded placebo controlled RCT</td>
</tr>
<tr>
<td>Immunotherapy-naive (%)</td>
<td>100%</td>
<td>100%</td>
<td>53%$^{a}$</td>
</tr>
<tr>
<td>Nephrectomized (%)</td>
<td>100%</td>
<td>100%</td>
<td>89%</td>
</tr>
<tr>
<td>Clear-cell histology (%)</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Female (%)</td>
<td>32%</td>
<td>27%</td>
<td>25%</td>
</tr>
<tr>
<td>Median age (range)</td>
<td>61 (30–82)</td>
<td>60 (18–81)</td>
<td>59 (28–85)</td>
</tr>
<tr>
<td>MSKCC risk score 0 (%)</td>
<td>27%</td>
<td>38%</td>
<td>39%</td>
</tr>
<tr>
<td>MSKCC risk score 1–2 (%)</td>
<td>56%</td>
<td>56%</td>
<td>55%</td>
</tr>
<tr>
<td>MSKCC risk score $\geq 3$ (%)</td>
<td>9%</td>
<td>6%</td>
<td>3%</td>
</tr>
<tr>
<td>MSKCC risk score NA (%)</td>
<td>9%</td>
<td>0%</td>
<td>3%</td>
</tr>
</tbody>
</table>

Study arms

<table>
<thead>
<tr>
<th>Tumor assessment</th>
<th>RECIST</th>
<th>RECIST</th>
<th>RECIST</th>
</tr>
</thead>
<tbody>
<tr>
<td>BEV+IFN</td>
<td>IFN</td>
<td>IFN</td>
<td>IFN</td>
</tr>
<tr>
<td>SUN</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PAZ</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: $^{a}$Patient characteristics refer to the overall population including therapy-naïve and pretreated patients however the PFS outcomes used in the ITC are based on the first-line (treatment-naïve) subpopulation.

Abbreviations: AVOREN, AVastin fOr RENal cell cancer; BEV, bevacizumab; IFN, interferon-α-2a; MSKCC, Memorial Sloan-Kettering Cancer Center; NA, not available; PAZ, pazopanib; PFS, progression-free survival; PLA, placebo; RCT, randomized controlled trial; RECIST, Response Evaluation Criteria in Solid Tumors; SUN, sunitinib.

Figure 2 Indirect treatment comparison: efficacy connections between the pivotal trials.

Abbreviations: AVOREN, AVastin fOr RENal cell cancer; BEV, bevacizumab; IFN, interferon-α-2a; PLA, placebo; PAZ, pazopanib; RCT, randomized controlled trial.
For both scenarios the indirect comparison HR of BEV+IFN vs TKIs with 95% CIs are provided. Figure 3 shows the detailed calculation pathway defined by Bucher et al, including the BEV+IFN vs SUN comparison.

For the comparison of BEV+IFN vs PAZ the same methodology was applied, but two ITCs needed to be performed in contrast to the SUN comparison. In a first step, the ITC HR of PAZ vs IFN was calculated (using the published PAZ PFS HR and the connector trials’ PFS HR) and in a second step this ITC HR result was compared to the published PFS HR of BEV+IFN.

All calculations have been performed in Excel 2003 (Microsoft, Redmond, WA). The ITC calculations can be reperformed using the ITC tool available from the Canadian Agency for Drugs and Technologies in Health, which ensures maximum transparency.

**Patient compliance**

As the TKIs, SUN and PAZ are oral medications that are self-administered by the patient and show a considerable adverse event profile, compliance effects are expected in real-world settings.

---

**Table 2 Overview of selected connector trials**

<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Comparison</th>
<th>Total N</th>
<th>Median PFS in months</th>
<th>PFS HR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MRCRCC trial</td>
<td>1999</td>
<td>IFN vs MPA</td>
<td>335</td>
<td>4.0</td>
<td>0.72 (0.56–0.92)</td>
</tr>
<tr>
<td>Aass et al</td>
<td>2005</td>
<td>IFN plus 13-CRA vs IFN</td>
<td>320</td>
<td>5.1</td>
<td>0.66 (0.52–0.85)</td>
</tr>
<tr>
<td>Mickisch et al</td>
<td>2001</td>
<td>IFN plus nephrectomy vs IFN</td>
<td>84</td>
<td>5.8&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.60 (0.36–0.97)</td>
</tr>
<tr>
<td>Proxy comparison</td>
<td>2009</td>
<td>IFN vs PLA</td>
<td>NA</td>
<td>5.4</td>
<td>0.52&lt;sup&gt;c&lt;/sup&gt; (0.32–0.72)</td>
</tr>
</tbody>
</table>

**Notes:** Estimated median PFS for IFN and PLA, respectively; Median time to progression; Estimated HR based on Phase III median PFS outcomes assuming constant hazards.

**Abbreviations:** 13-CRA, 13-cis-retinoic acid; HR, hazard ratio; IFN, interferon-α-2a; MPA, medroxy-progesterone acetate; MRCRCC, Medical Research Council Renal Cancer Collaborators; NA, not available; PLA, placebo; PFS, progression-free survival.

---

**Figure 3** Indirect comparison methodology according to Bucher et al showing the calculations for the comparison of BEV+IFN vs SUN.

**Notes:** Starting with the published PFS HRs, study-specific log-hazard ratios are calculated (step 1) and transferred into an ITC HR (step 2); The indirect treatment comparison 95% confidence intervals are calculated using the 95% CI of the published PFS HR (steps 3–5).

**Abbreviations:** BEV, bevacizumab; IFN, interferon-α-2a; ITC, indirect treatment comparison; SUN, sunitinib.
In order to estimate patient compliance under routine conditions, data obtained from IMS Health were used as the basis for the estimation. These data were obtained on the basis of 1869 Dutch mRCC patients treated with sunitinib. Physician records have been used in order to determine patient compliance at different points in time. According to these data the median SUN compliance rate at 3, 6, and 9 months of SUN therapy was 74%, 72%, and 71%, respectively.

As there are currently no published data on PAZ patient compliance available, it was assumed that the compliance rates are comparable to SUN. We performed analyses using the conservative estimates of 90%, 80%, and 70% patient compliance for the TKIs, respectively.

In order to simulate the compliance impact on the PFS of the TKIs an adjustment of the PFS HR was performed. As no clinical trial data are available that show the effectiveness impact of noncompliance, it was conservatively assumed that the PFS HR of noncompliant patients is ‘1’, which means the same efficacy as for IFN, and that the published Phase III efficacy refers to the compliant patients.

The detailed steps taken to estimate the real-world effectiveness (adjusted TKI PFS HR), depending on patient compliance, are shown in Figure 4, using a 70% SUN patient compliance as an example. The same approach was applied for all scenarios analyzed.

As BEV is infused intravenously, the patient either visits the physician to receive the injection or decides to stop therapy by not attending. Even though many patients self-administer the subcutaneous IFN injections (in combination with BEV), which might be a potential compliance issue, downdosing of IFN has been shown to improve tolerability and maintain efficacy.21 Hence it was assumed that missing an IFN injection has a limited impact on the PFS HR of BEV+IFN, so no patient compliance impact on BEV+IFN therapy was simulated.

**Results**

The indirect efficacy comparison, shown in Figure 5, resulted in a statistically nonsignificant PFS difference of BEV+IFN vs SUN (ITC HR: 1.06; 95% CI: 0.78–1.45; \( P = 0.73 \)) and of BEV+IFN vs PAZ (range based on different connector trials; ITC HR: 0.74–1.03; \( P = 0.34–0.92 \)).

For the BEV+IFN vs PAZ comparison the two extreme scenarios are based on the selected connector trials, whereby using the MRCRCC trial resulted in an ITC HR of 1.03 (95% CI: 0.61–1.74; \( P = 0.92 \)) and using the ‘proxy comparison’ resulted in an ITC HR of 0.74 (95% CI: 0.40–1.37; \( P = 0.34 \)).

Simulating real-life patient compliance and its effectiveness impact on PFS showed an increased tendency towards BEV+IFN without reaching statistical significance, as shown in Figure 6.

For the comparison of BEV+IFN vs PAZ simulations have been performed for the extreme scenarios, which means the connector trials producing the highest ITC HR (MRCRCC Trial) and the lowest ITC HR (proxy comparison) have been analyzed.

**Discussion**

Comparing the PFS efficacy and effectiveness of BEV+IFN vs the TKIs SUN and PAZ in first-line mRCC therapy failed to show a significant tendency in favor of one particular targeted therapy approach. Additionally, the influence of

![Figure 4 Patient compliance PFS adjustment methodology.](image-url)

**Abbreviations:** HR, hazard ratio; IFN, interferon-\( \alpha \)-2a; PFS, progression-free survival; SUN, sunitinib.
Patients compliance on the PFS was investigated. This indirect effectiveness assessment indicates that the PFS outcomes with regard to TKIs might be lower in real-world settings. However, the observed tendency towards a better effectiveness of BEV + IFN failed to reach statistical significance.

The main limitation is that our findings are based on indirect evidence. Such an indirect treatment comparison has to be regarded as a complementary assessment to clinical trials, because it cannot substitute direct evidence. However, in the absence of any head-to-head comparison, the indirect treatment comparison approach should be regarded as the most valuable way of estimating treatment effects in a statistically accurate manner.

Another limitation is that there is no matching connector trial available in order to determine an exact ITC hazard ratio for the comparison of BEV + IFN vs PAZ. The lack of an adequate connector trial, comparing IFN vs PLA, was overcome by using different but the most suitable IFN studies in order to enable a bridge to be built between the PAZ and the BEV + IFN PFS outcomes. Furthermore, an additional ‘proxy comparison’ was performed that is based on assuming constant hazards to estimate a HR of IFN vs PLA based on the available Phase III evidence. The authors would like to point out that the application of constant hazards should be performed very carefully but in this special case (no adequate connector trial available) it was decided to perform this analysis to test the credibility of the bridging trials’ HR on the ITC results. As no statistically significant difference was observed when comparing the PFS HR of BEV + IFN vs PAZ, irrespective of the connector trial used, the lack of an adequate bridging trial is regarded as having a limited impact on the ITC results.

Another limitation is that data on patient compliance to TKIs are currently rare. We used IMS data, which refers to the Dutch health care system, to estimate the proportion of patients who show a limited compliance to TKI therapy. As there are no real world investigations available that determine the impact of patient compliance on the PFS, we used a conservative assumption. However, further research is required in order to evaluate a more accurate link between patient compliance and its impact on efficacy.

Another aspect to be considered is the difference in patient characteristics between the pivotal trials used. According to the patient’s risk profile, the AVOREN study included fewer patients with a favorable disease prognosis; hence the PFS outcomes might be underestimated in comparison to SUN and PAZ. However, as all patients have been nephrectomized in the AVOREN trial, which is regarded as an indicator for a better disease prognosis, these small differences in prognostic patient characteristics are estimated to compensate each other.

In the past there was a consensus that SUN and BEV + IFN are equally effective in terms of PFS in first-line mRCC therapy, which is in line with our findings. However, recent publications raised doubts about this comparable efficacy. Both papers focused only on investigator-assessed PFS values and pooled BEV + IFN PFS outcomes from a strictly controlled pivotal Phase III trial and an

**Table 1: Indirect efficacy comparison results PFS HR of BEV + IFN vs TKIs.**

<table>
<thead>
<tr>
<th>Connector Trial</th>
<th>BEV + IFN vs PAZ</th>
<th>BEV + IFN vs SUN</th>
</tr>
</thead>
<tbody>
<tr>
<td>MRCRCC</td>
<td>1.03 (0.61 – 1.74)</td>
<td>1.06 (0.78 – 1.45)</td>
</tr>
<tr>
<td>Aass et al</td>
<td>0.94 (0.56 – 1.59)</td>
<td>0.86 (0.44 – 1.69)</td>
</tr>
<tr>
<td>Mickisch et al</td>
<td>0.86 (0.44 – 1.69)</td>
<td>0.74 (0.40 – 1.37)</td>
</tr>
<tr>
<td>Proxy comparison</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 5** Indirect efficacy comparison results PFS HR of BEV + IFN vs TKIs.

**Abbreviations**: BEV, bevacizumab; IFN, interferon-α-2a; TKI, tyrosine kinase inhibitor; CI, confidence interval; HR, hazard ratio; SUN, sunitinib.
investigator-initiated trial. As a result of this pooling, the efficacy of BEV+IFN was decreased on the basis of a lower PFS observed in the investigator-initiated trial compared to the pivotal trial outcomes.

In order to ensure comparability, it was expected that the authors would apply the same procedure for SUN, using the pivotal trial and the first-line outcomes from the SUN expanded-access-study, but only the pivotal trial outcomes were used for SUN.

An adequate indirect comparison approach should take into account pivotal trials performed under the same conditions to be comparable and use the highest quality data (independent radiology review of PFS). Hence our approach focused on the comparison of the pivotal Phase III trials, using the highest data quality, in order to ensure comparability of therapy outcomes.

Our findings have been confirmed by another recently published indirect treatment comparison performed from the perspective of PAZ. McCann et al concluded “that pazopanib demonstrates no reduction in efficacy compared to other approved angiogenesis inhibitors”, which is in line with our findings that say ‘there is no significant difference in first-line PFS outcomes between BEV+IFN and the TKIs SUN and PAZ’. As a consequence there is a need for other...
clinical decision criteria that might allow an adequate therapy selection in first-line mRCC patients. Possible guidance might be offered by aspects of available therapy sequencing options, sequential therapy outcomes, and by tolerability issues.29

For example there is evidence that BEV+IFN shows a better tolerability profile if indirectly compared to SUN, which also impacts the costs of managing side effects.30,31 In addition, there are first retrospective analyses indicating that BEV+IFN first-line enables effective subsequent TKI therapy,6,32 which may lead to improved patient outcomes, taking into account the complete sequence of mRCC therapies.33,34

Conclusions
In conclusion, in the light of the currently available evidence, there is no statistically significant PFS difference between BEV+IFN and TKIs in first-line mRCC therapy. In terms of patient compliance there is an efficacy tendency in favor of BEV+IFN, but this fails to reach statistical significance.

These findings imply that other treatment decision criteria such as tolerability and therapy sequencing opportunities need to be considered in order to guide adequate therapy decisions.

Disclosure
This work was funded by F Hoffmann-La Roche Pharmaceuticals AG.

References


