Relative effectiveness of sunitinib versus everolimus in advanced pancreatic neuroendocrine tumors: an updated matching-adjusted indirect comparison

DOI:
[10.2217/cer-2018-0020](https://doi.org/10.2217/cer-2018-0020)

Document Version

Accepted author manuscript

[Link to publication record in Manchester Research Explorer](#)

Citation for published version (APA):

Ishak, K. J., Rael, M., Hicks, M., Mittal, S., Eatock, M., & Valle, J. W. (2018). Relative effectiveness of sunitinib versus everolimus in advanced pancreatic neuroendocrine tumors: an updated matching-adjusted indirect comparison. *Journal of Comparative Effectiveness Research*, 7(10), 947-958. <https://doi.org/10.2217/cer-2018-0020>

Published in:

Journal of Comparative Effectiveness Research

Citing this paper

Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

General rights

Copyright and moral rights for the publications made accessible in the Research Explorer are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Takedown policy

If you believe that this document breaches copyright please refer to the University of Manchester's Takedown Procedures [<http://man.ac.uk/04Y6Bo>] or contact uml.scholarlycommunications@manchester.ac.uk providing relevant details, so we can investigate your claim.

**Relative Effectiveness of Sunitinib vs. Everolimus in
Advanced Pancreatic Neuroendocrine Tumors: An Updated
MAIC**

Journal:	<i>Journal of Comparative Effectiveness Research</i>
Manuscript ID	CER-2018-0020.R2
Manuscript Type:	Research Article
Keywords:	sunitinib, everolimus, comparative effectiveness

SCHOLARONE™
Manuscripts

MAIC of Sunitinib vs. Everolimus in panNET

Relative Effectiveness of Sunitinib vs. Everolimus in Advanced Pancreatic Neuroendocrine Tumors: An Updated Matching-adjusted Indirect Comparison

Abstract:

Background: A Matching-adjusted Indirect Comparison (MAIC) of sunitinib and everolimus has been previously reported based on the RADIANT-3 everolimus trial. We performed an analysis using updated OS data based on sunitinib’s trial (A6181111).

Methods: The MAIC matched on all baseline characteristics available from both studies. An anchored MAIC was performed for PFS; an unanchored analysis was deemed more appropriate for OS due to crossover in both trials. A hazard ratio for sunitinib vs. everolimus was derived from adjusted (weighted) sunitinib effects compared to the observed results for everolimus.

Results: The adjusted hazard ratio for sunitinib vs. everolimus was 0.85 (0.39–1.89) for PFS and 0.82 (0.53–1.27) for OS.

Conclusions: Findings indicate comparable PFS and OS with sunitinib and everolimus.

Keywords: sunitinib, everolimus, matching-adjusted indirect comparison (MAIC), comparative effectiveness

Background

Neuroendocrine neoplasms of the pancreas account for a minority (<5%) of pancreatic malignancies; their clinical behavior, driven by their pathology, ranges from aggressive malignancy (poorly differentiated, grade 3 neuroendocrine carcinomas) to a more indolent spectrum (well-differentiated [grade 1 and 2] neuroendocrine tumors [NETs]).[1] Although a small proportion of patients with pancreatic NET harbor an inherited predisposition, most are sporadic. Additionally, most are non-functional, although some patients have discrete clinical syndromes due to hormonal hypersecretion (e.g., insulinomas, glucagonomas, etc.).[2]

Surgery for patients with early-stage disease is the cornerstone of therapy for patients with well-differentiated NETs; in advanced disease, chemotherapy (with older agents such as streptozocin, 5-fluorouracil and doxorubicin) had been used for decades in the absence of new therapies.[3] More recently, both sunitinib and everolimus have been approved for the treatment of patients with advanced, progressive, well-differentiated pancreatic NETs (panNET), changing the treatment paradigm. Their approval was based on the evidence of efficacy in prolonging progression-free survival compared to best supportive care (BSC + placebo) shown in their respective Phase III trials (A6181111[4] compared BSC +

MAIC of Sunitinib vs. Everolimus in panNET

sunitinib vs. BSC + placebo, and RADIANT-3[5] compared BSC + everolimus vs. placebo). At the first analysis, the A6181111 trial also suggested an overall survival benefit, although this failed to reach statistical significance, most likely due to the small sample size and the confounding effect of crossover from the BSC + placebo arm to sunitinib.

Given the simultaneous emergence of these two therapies (both published in 2011), there are no head-to-head studies comparing sunitinib and everolimus; therefore, evidence of their relative effectiveness needs to be derived through indirect comparisons of outcomes. Typically, this would be handled using a network meta-analysis (NMA)[6-8] since the trials share common reference treatments. A NMA may be unreliable, however, when the design and/or population of the studies in the network differ in a significant way. A specific challenge in the context of the sunitinib and everolimus trials is crossover from placebo to respective active drug. That is, patients who progress with BSC + placebo could cross over and receive sunitinib or everolimus in an open-label, extended follow-up phase of the studies. Although this has no impact on the assessment of progression-free survival (PFS) (since progression precedes crossover), comparison of overall survival (OS) is distorted, as outcomes in the BSC + placebo arms reflect a mixture of the effects of BSC and sunitinib or everolimus after progression. If these treatments have a differential effect on OS, the BSC + placebo arms of the two trials cannot be considered comparable. Furthermore, crossover may have occurred differently in the two trials in terms of the extent and timing of crossover, as well as the characteristics of patients who cross over. Thus, a NMA, which would rely on the BSC + placebo arms as an anchor for the comparison, may be unreliable as the arms may not be comparable across the trials due to the effects of crossover.

An indirect comparison is possible without an anchor arm by utilizing techniques such as a matching-adjusted indirect comparison (MAIC).[9] A MAIC was previously performed for sunitinib and everolimus based on the A6181111[4] trial (sunitinib vs. BSC + placebo) and the RADIANT-3 trial[5] for everolimus.[10] The authors concluded that both PFS and OS were similar for the two treatments. More mature OS data have become available for both sunitinib[11] and everolimus[12] trials since the publication of the aforementioned MAIC; the data provide additional follow-up for OS and allow for the incorporation of additional baseline variables in the matching. An updated MAIC of sunitinib and everolimus was, therefore, carried out to assess relative OS using the data that included longer follow-up.

Methods

Overview of the A6181111 and RADIANT-3 Trials

The A6181111 and RADIANT-3 studies are both Phase III randomized trials in patients with advanced, progressive, well-differentiated panNETs. Both studies had a similar design (placebo control, PFS as the primary endpoint and BSC in the control arms). Although the sample sizes in the studies were comparable

MAIC of Sunitinib vs. Everolimus in panNET

based on the statistical design of each study, the studies differed in terms of final recruited population size, with A6181111 recruiting 171 patients (of 340 planned; with early closure on the advice of the Independent Safety Monitoring Committee) and RADIANT-3 recruited 410 patients, as planned. The inclusion and exclusion criteria were generally similar between the trials; both recruited patients with well-differentiated (grade 1 and 2) panNETs with radiological progression documented within the prior 12 months. A notable difference between the studies exists in the performance status (PS) of patients at enrollment; A6181111 included patients with a PS score of 0 or 1, while RADIANT-3 allowed patients with a score of 2. Only 3% of patients had a score of 2, however, which is likely to have had only a slight impact, if any, on the results from the trial. The characteristics of the two populations are summarized in Table 1.

The A6181111 study terminated early on the recommendation of the safety monitoring committee due to “the greater number of deaths and serious adverse events in the placebo group and the difference in progression-free survival favor of sunitinib”[4] at an interim analysis of the data. As a result, follow-up for PFS was shorter in A6181111, although OS follow-up continued beyond the end of both trials. The studies also differed with respect to the schedule for radiological assessment based on interpretation of CT scans in accordance with RECIST evaluation, which was every 8 weeks in A6181111 and every 12 weeks in RADIANT-3. This implies that progression may have been detected earlier in A6181111.

Both studies demonstrated PFS gain when either sunitinib or everolimus was initiated vs. BSC + placebo, with hazard ratios (HR) indicating an approximate 60% reduction in risk (0.42 (0.26–0.66) for sunitinib and 0.35 (0.27–0.45) for everolimus). Crossover from the BSC + placebo arm to the experimental arm occurred in both studies. In A6181111, 69% of patients in the BSC + placebo group crossed over at progression or early termination of the trial. In RADIANT-3, 85% of the reference group had crossed over at progression at the latest data cut.[12] The HRs for OS from the latest data cuts were 0.73 (0.50–1.06) for sunitinib and 0.94 (0.73–1.20) for everolimus.

Statistical Methods

Overview of Analytic Approaches

The primary approach for indirect comparison was based on a MAIC; a traditional comparison using Bucher’s approach³ was also applied as a reference. This was derived as the ratio of the HRs for PFS and OS in each of the trials, with confidence intervals reflecting uncertainty from both results. That is, the standard error (SE) of the log of the ratio of HRs was calculated as the sum of the SEs of each of the study-specific log HRs.

The MAIC analyses used patient-level data from A6181111 (index trial) and aggregate data reported in publications for the comparator trial (RADIANT-3).[5, 12] The A6181111 data are analyzed to derive weights for each of the patients in the trial—this yields weighted average values of baseline characteristics

MAIC of Sunitinib vs. Everolimus in panNET

(which include proportions for categorical variables) that match those of the RADIANT-3 population. The weights were derived from a propensity score–type equation (logistic regression) that predicts the odds that a patient is from the A6181111 or the RADIANT-3 trial. Since only aggregate data are available for the comparator trial, the coefficients of the equation must be estimated using the method of moments.[9]

Success of matching was verified by comparing weighted mean characteristics of the index groups to match their target values. The effective sample size (ESS¹) of the index population after weighting was calculated to reflect the number of patients contributing to calculations where the weights are used. The ESS is determined by the extent of differences between the populations being matched—the larger the discrepancy between the populations, the lower the ESS. The ESS was also used to adjust the standard errors and confidence intervals (CI) of measures of relative effectiveness so that these reflect the actual extent of data contributing to the analyses. This is done by normalizing the weights to add up to the ESS for the group; as these weights are applied in calculations of weighted outcomes, standard errors are also adjusted accordingly to reflect the ESS rather than the original sample size.

The balancing weights were applied to derive adjusted measures of outcomes with the index treatments in the comparators population. Two possible approaches are possible: one is an anchor-based analysis, which can be used when the treatments used in the reference arms are the same in the two trials. In this approach, the effect within the index trial (i.e., sunitinib vs. BSC + placebo) is recalculated after applying the balancing weights to the two arms. This adjusts the HR to the characteristics of the comparator's population while preserving randomization within the index trial. A Bucher-type analysis was then applied to the adjusted HR and the reported HR for the comparator to obtain an estimate of relative effectiveness. An anchor-based analysis is possible for PFS in this MAIC, but not for OS since the BSC + placebo arms are contaminated by crossover. The second approach, which involves an arm-to-arm comparison after matching, is necessary to compare OS. In this case, weights were applied to sunitinib patients to derive an adjusted OS distribution (or Kaplan–Meier curve), which can be compared with the OS distribution of everolimus to derive an HR. The specific approach used in these analyses, including the arms being matched and how balancing weights were applied to derive adjusted outcomes, are described in the sections below.

The key features and assumptions of each approach and their appropriateness in this analysis are summarized in Table 1. The Bucher approach relies on the key assumption that the common BSC arms of the studies are comparable, and that there are no effect modifiers (i.e., subgroups within which the effect of sunitinib or everolimus varies). The anchored MAIC relaxes this requirement by matching on baseline

¹ The ESS is given by $(\sum w_i)^2 / (\sum w_i^2)$, where w_i are the individual weights derived for patients in the index trial.

MAIC of Sunitinib vs. Everolimus in panNET

characteristics, assuming that all effect modifiers are captured in the matching. The unanchored MAIC overcomes any issues of comparability of the BSC arms by deriving effects directly from the sunitinib and everolimus arms after adjustment for potential predictors and effect modifiers. By doing so, this approach breaks randomization and is susceptible to residual confounding if some predictors or effect modifiers are not included in the matching. Both MAIC-based analyses adjust to the population of RADIANT-3 and thus invoke an added assumption that the derived relative effects are applicable to the sunitinib population, as this is the basis of the comparison.

Variables Included in Matching

All commonly available variables from A6181111 and RADIANT-3 were included in the MAIC to ensure as close a match of the populations as possible. More specifically, the following variables were included: median age; the proportion >64 years; gender; race (white vs. other); Eastern Cooperative Oncology Group (ECOG) performance status; proportion of time since diagnosis of more than three years; number of disease sites (1, 2, 3+); presence of metastases in the liver, lymph nodes, and lung; prior use of somatostatin analogs; concurrent use of somatostatin analogs; and prior treatment with systemic chemotherapy or prior radiotherapy treatment. Although arm-specific characteristics were available for almost all variables, the distribution of race, time since diagnosis, and concurrent use of somatostatin analogues were only known for the combined population of the RADIANT-3 trial. These, therefore had to be used in matching; given randomization, it is assumed that these values are very close to the arm-specific values and using them for matching should not impact results.

Comparison of PFS

The MAIC of PFS was carried out as an anchor-based analyses; an arm-to-arm comparison was also performed to assess robustness of results.

For anchor-based analyses, weights were derived to match the sunitinib and BSC + placebo arms in A6181111 to the characteristics of the everolimus arm. Matching to a common target ensures the weighted groups are comparable. Sensitivity analyses were performed to match the two index trial arms to the BSC + placebo group in RADIANT-3 to rule out any influence from residual differences in the everolimus and reference groups (which is unlikely given randomization).

A Cox proportional hazards model was fitted to the A6181111 data using the normalized weights to match each of the two arms to the characteristics of the everolimus group. This produces the expected HR for sunitinib vs. BSC + placebo in a RADIANT-3-like population. The standard error for the HR is properly adjusted to the ESS after matching due to use of the normalized weights. The relative effectiveness of sunitinib vs. everolimus was then calculated as the ratio of the adjusted HR for sunitinib vs. BSC + placebo and the HR for everolimus vs. BSC + placebo reported in RADIANT-3 (0.35 [0.27–0.45]). A 95% CI was derived for this estimate using the square root of the sum of the variances of the two log-HRs used in the

MAIC of Sunitinib vs. Everolimus in panNET

calculation. The variance of the adjusted HR was produced directly from the analyses, while the variance of the HR for everolimus was approximated from its 95% CI: $\log[\text{Upper Bound/Point Estimate}]/1.96$.

As crossover does not impact on PFS, a standard Bucher calculation was applied on the published HRs for each of the treatments in their respective trials. This represents an unweighted analysis and served as a reference for MAIC-based calculations. An arm-to-arm comparison of PFS between sunitinib vs. everolimus was also carried out; this used balancing weights to match the two arms in the comparison and followed the approach described in the next section for OS.

Comparison of OS

For OS comparisons, updated survival data for everolimus were taken from a poster presented at the American Society of Clinical Oncology (ASCO) conference in 2015[12], and the most recent data cut of the A6181111 trial was used for OS with sunitinib. Since the BSC + placebo arms are contaminated by crossover, an arm-to-arm comparison was carried out between sunitinib and everolimus to compare OS. These analyses used weights derived from matching the sunitinib arm to the characteristics of the everolimus arm.

The Kaplan–Meier curve for OS for everolimus was digitized using ENGAUGE software, and converted to virtual patient-level (VPL) data using the approach of Hoyle and Henley.[13] This produces a sample of the same size as the everolimus group, with individuals assigned a time of death or censoring with an appropriate indicator based on the approximated counts of events and patients at risk.[13] These VPL data for everolimus are assigned a weight of 1 and combined with the patients from the sunitinib group who are assigned the normalized weights derived from the matching step.

A Cox proportional hazards model is then fitted to the index and VPL data, including an indicator for treatment group for sunitinib vs. everolimus to capture the relative effect of sunitinib and everolimus. This is expressed as an HR with a 95% CI.

Residual Confounding and Potential Study Effects

MAICs rely on the assumption that all potential confounders are included in the matching and that other differences between studies (i.e., study effects) do not further distort comparisons. Anchor-based analyses, as described above for PFS, inherently account for this by using a relative effect estimate as the basis for comparison. This is not the case in arm-to-arm comparisons (as for OS), which are susceptible to potential biases due to missing matching variables or other differences. When the trials have common reference treatment arms, this assumption can be assessed by matching the reference arms and assessing any differences in outcomes. Similar results in the common arms after matching would support the validity of the results.

MAIC of Sunitinib vs. Everolimus in panNET

The BSC + placebo arms in the trials in this MAIC can be used to assess potential biases (due to residual confounding and study effects) in arm-to-arm comparisons. Given contamination by crossover, such an assessment is not possible based on OS in the BSC + placebo arms. It was, therefore, executed based on PFS, with the caveat that observed differences may not necessarily extrapolate to a bias in OS. For instance, imaging in A6181111 was carried out every 8 weeks, while RADIANT-3 followed a 12-week interval. This affects the timing of detection of progression, and as a result, affects the comparison of PFS curves, but it would not affect OS since that is assessed based on actual dates of death.

Balancing weights were derived for the BSC + placebo arm in A6181111, matching it directly to the BSC + placebo arm for RADIANT-3. An HR quantifying differences in PFS was derived following the approach described for OS analyses. This involved converting the RADIANT-3 PFS curve to VPL, combining these data with the weighted patient-level data from A6181111, and applying a Cox model to obtain the adjusted HR.

Results

Matching of Populations

Table 2 summarizes the baseline characteristics of the two arms of the sunitinib trial before and after matching to the everolimus arm in RADIANT-3. The most notable differences between the populations are in terms of race (the proportion of white patients is lower in A6181111), ECOG = 0 in the BSC + placebo arm (higher in A6181111), number of metastatic sites (lower in A6181111), concurrent use of somatostatin analog use (lower in A6181111), prior use of systemic chemotherapy (higher in A6181111), and prior use of radiotherapy (lower in A6181111). After matching, the populations have identical values for these characteristics.

The table also highlights some notable differences between the arms in the A6181111 trial itself, despite randomization. In particular, the sunitinib and BSC + placebo arms differed at baseline in terms of ECOG PS, time since diagnosis, and metastasis sites. This implies that comparisons within the trial are susceptible to confounding by these differences.

The ESS after matching was 43 in the sunitinib arm and 31 in the BSC + placebo arm. The smaller ESS in the BSC + placebo arm is due to the larger differences in the proportion of patents with ECOG PS = 0, disease duration, and metastasis of lymph nodes. Matching was also done between the two BSC + placebo arms to assess residual confounding, which led to an ESS of 37.

Comparison of PFS

Both sunitinib and everolimus demonstrated improvement in PFS, with trials showing HRs in the order of 0.40 and reaching statistical significance (Table 3). A Bucher analysis applied to published results yields a

MAIC of Sunitinib vs. Everolimus in panNET

relative HR of 1.20 (0.72–2.01), indicating no statistically significant difference between sunitinib and everolimus.

The weighted HR for sunitinib vs. BSC + placebo for the anchor-based MAIC was 0.30 (0.14–0.63), which implies a stronger effect than the unweighted HR (0.42 [0.26–0.66]). A comparison with the HR observed in RADIANT-3 for everolimus vs. BSC + placebo (0.35 [0.27–0.45]), yields a relative HR of 0.85 (0.39–1.89) for sunitinib vs. everolimus.

An arm-to-arm comparison of PFS was also carried out in sensitivity analyses—Figure 1 illustrates the observed PFS curves for sunitinib and everolimus as well as the sunitinib curve after weighting to the everolimus curve. Comparison of the sunitinib-matched and everolimus curves yielded an HR of 0.86 (0.48–1.55), which is consistent with anchor-based analyses.

Comparison of OS

The OS data used in these analyses included 55 deaths in the sunitinib group, representing 64% of patients in the arm, and 126 deaths (60.9%) for everolimus. Figure 2 shows the observed OS curves based on the most recently available data cuts for sunitinib and everolimus, as well as the curve obtained from the everolimus-matched population for sunitinib. The effect of matching to the everolimus population profile is most apparent after around month 20, where the matched curve separates from the original (unmatched OS curve) and is characterized by relatively long flat periods. This reflects the combined effect of the drop in ESS after matching and greater weight being assigned to long-term survivors with sunitinib after matching.

A combined analysis of the patient-level data for sunitinib and the VPL for everolimus produced an HR of 1.03 (0.75–1.42) when the populations were not matched. Applying the balancing weights (and hence matching the populations) yields an HR of 0.82 (0.53–1.27)—the median survival with sunitinib after matching was 53.4 months (38.6 months unmatched) and 44.0 months with everolimus.

Residual Confounding and Study Effects (Based on PFS)

Residual confounding (i.e., unbalanced differences between populations) or bias due to other differences between the studies (i.e., study effects) was assessed by comparing the BSC + placebo arms of the A6181111 and RADIANT-3 trials (Figure 3). Although some early separation can be seen between the curves, this may be due to differences in the assessment cycles in the two trials. The curves otherwise overlap closely, indicating similar PFS with BSC + placebo in the two trials after matching. This is confirmed by the HR derived from comparing these curves (0.98 [0.62–1.54]).

MAIC of Sunitinib vs. Everolimus in panNET

Discussion

The A6181111 and RADIANT-3 trials have demonstrated similar clinical benefit with a reduction of 58% and 65% in the risk of progression or death with sunitinib and everolimus, respectively, compared with BSC. This MAIC of sunitinib and everolimus in patients with well-differentiated panNETs showed comparable PFS and OS, with relative effect estimates (i.e., HRs) that were statistically significant (since the confidence intervals included the null HR=1). This may be due in part to the relatively small sample size of the sunitinib trial and the further reduction in ESS after matching, which can limit the power to detect significance. A comparison of PFS in the BSC + placebo curves of the two trials after matching suggested very similar outcomes, with a HR of 0.98. This supports the reliability of these analyses, as it suggests that there is no evidence of residual confounding or study effects that may distort the results for comparisons of sunitinib and everolimus.

Our results agree with those from the prior MAIC by Signorovitch and colleagues,[10] which found sunitinib and everolimus to be comparable in terms of PFS and OS. The methods in the two analyses followed the same general approaches. A difference of note is in the variables included in matching; this MAIC included all the same variables used previously and supplemented them with race, prior radiotherapy treatment, and concurrent use of somatostatin analogues, and replaced presence of distant metastases with specific sites of metastases (liver, pancreas, lymph node, lung). The aim was to ensure a close match between the populations being compared. The HR for PFS for everolimus vs. BSC + placebo used by Signorovitch and colleagues differed from the result reported by Yao and colleagues[5] in the trial publication (0.38 vs. 0.35)—this may have been due to exclusion of observations due to missing data in baseline characteristics.

Although the general conclusions from the analyses agreed, some differences were noted in the point estimates obtained. For instance, both yielded confidence intervals suggesting no statistically significant differences between treatments, but the point estimates went in opposite directions, with the current analyses favoring sunitinib. For PFS, this MAIC yielded an HR of 0.85 (0.39–1.89) for sunitinib vs. everolimus, while the prior analyses found an HR of 1.20 (0.65–2.17)[10] for the same comparison. The reversal of direction of the estimate is due to a notable change in the sunitinib vs. BSC + placebo HR, which went from 0.42 to 0.30 after matching. This change in HR is because of the baseline imbalances between the sunitinib and BSC + placebo arms in the A6181111 trial noted above—in particular, in the proportion of patients with ECOG PS = 0, disease duration of three or more years, and number of sites. Since both groups are matched to the characteristics of the same comparator population, the weighting of patients in the calculation of the HR effectively balances out the baseline differences (in addition to matching the groups to the comparator population). Thus, if the prior MAIC[10] had used an adjusted HR for sunitinib vs. BSC + placebo, results should have aligned in direction. To verify this, we derived the HR for sunitinib vs. BSC + placebo after adjusting for baseline variables included in the MAIC; this gave an HR of 0.32 (0.19–0.53),

MAIC of Sunitinib vs. Everolimus in panNET

which contrasted with the weighted HR for everolimus vs. BSC + placebo, which yielded a relative HR 0.91 (0.48–1.74) for sunitinib vs. everolimus.

The direction of the point estimates from OS analyses also differed in a similar way, suggesting a potential signal of improved survival with sunitinib (HR: 0.82 [0.53–1.27]) vs. everolimus, while the analyses by Signorovitch and colleagues[10] yielded an HR of 1.23 (0.76–2.04). This may be due to changes in survival in the more recent cuts of OS used in the current analyses, or possibly the fact that more variables are included in the matching. The latter was assessed by repeating analyses using the same matching variables as those used in the prior MAIC; the HR derived from these analyses was 0.80 (0.54–1.19), which is nearly identical to the one obtained with all variables. Therefore, the results are likely reflective of changes in the updated OS data.

The direction of HR derived in these analyses appears to be mostly driven by a separation of the OS curves after 30 months. This may reflect a delayed benefit for sunitinib, but may also be spurious due to the relatively small sample size in the sunitinib group (particularly after matching). Results may also be impacted by differences in long-term treatment choices between the groups. This is not factored into the current analyses and may influence the shape of the curves over the longer term. Therefore, relying on the earlier OS patterns where the data are richer provides a conservative assessment of the analyses that would conclude comparable survival.

The key limitation in these analyses is the small sample size of the sunitinib trial. Although more mature data on OS were available for these analyses, the sample provided only 55 deaths. This limits the possibility of detecting significant differences, as well as of exploring and capturing time-varying effects between the arms. Analyses matched on a broad range of variables; however, some minor differences remained. For instance, the RADIANT-3 trial included patients with ECOG level 2, while A6181111 excluded these patients; it is therefore impossible to balance out this difference. Given that this only represents 3% of patients in RADIANT-3, its impact is deemed to be minimal. Similarly, two patients in the sunitinib trial had no metastases, while all in RADIANT-3 had at least one. Given the very low count, this difference is also considered inconsequential.

The studies may have also differed in terms of other factors that were not available from the two trials, like the presence of syndrome and the presence of inherited genetic conditions. Tumor differentiation is another potential source of residual confounding; all patients in A6181111 had well-differentiated tumors based on the World Health Organization classification, of which 18% corresponded to the well-differentiated definition used in RADIANT-3 (based on the American classification), and 82% would be considered moderately differentiated. The composition of the RADIANT-3 population was predominantly well differentiated (83%). Given the large discrepancy between studies, adjustment for this factor would lead to significant reductions in the effective sample size after matching. The PFS comparisons in our analyses should not be affected as these were anchored to the BSC arms; OS analyses are subject to

MAIC of Sunitinib vs. Everolimus in panNET

residual confounding, however. We believe the impact may be limited since subgroup analyses in this trial [5] did not detect any differential effect in subgroups; furthermore, our assessment of residual confounding based on PFS in the BSC arms did not reveal any significant differences in outcomes after matching on other factors. Another limitation is the role of subsequent therapies in each arm that may have impacted overall survival (including chemotherapy, somatostatin analogues, peptide receptor radionuclide therapy, etc.). Matching included prior and concurrent use of somatostatins; however, subsequent treatment options may vary depending on patient-specific factors, but will also vary by region based on availability and reimbursement of each therapy option, but cannot be included in the matching to balance the comparison.

Sensitivity analyses were carried out to assess the robustness of results to assumptions. Anchor-based PFS analyses weighted the two arms of the A6181111 trial to the characteristics of the everolimus arm. We repeated the anchor-based MAIC matching the A6181111 groups to the characteristics of BSC + placebo arm of RADIANT-3 and obtained consistent results. The weighted HR for sunitinib vs. BSC + placebo was 0.32 (0.16–0.65) (compared with 0.30 [0.14–0.63]), and the indirect comparison of sunitinib vs. everolimus was 0.93 (0.44–1.94) (compared with 0.85 [0.39–1.89]). We also performed an arm-to-arm MAIC for PFS, as in the OS analyses. The HR obtained from this comparison was 0.86 (0.48–1.55), which is consistent with anchor-based analyses.

We also assessed whether the variables included in the matching affected the results; in particular, matching on a relatively large number of variables when the sample size of the index trial is relatively small was a potential concern as it may make the analyses susceptible to spurious results and limit generalizability. We repeated analyses matching only on variables previously used in the MAIC by Signorovitch and colleagues.[10] These produced very similar results to those from the full set of matched variables. For instance, the MAIC of PFS for sunitinib vs. everolimus was 0.95 (0.49–1.83) with the reduced set compared with 0.85 (0.39–1.89) when including all available variables. The difference may be indicative of additional adjustment provided by the variables added in the matching in the current analyses. Similarly, OS comparison with the reduced set of variables was 0.80 (0.54–1.19) compared with 0.82 (0.53–1.27) with all available variables matched. The effect of matching on additional variables can be seen on the width of the confidence intervals, however, which are wider when all variables are included.

The lack of an anchor for OS comparisons is another limitation of these analyses, as it forces breaking randomization when comparing across trials. This potentially leaves the comparisons susceptible to confounding due to variables that are not available for matching or differences in study design or execution that may have affected results. No specific issues that may cause such distortions were identified. Furthermore, an assessment of potential residual confounding based on PFS compared across the two BSC + placebo arms after matching showed nearly identical outcomes. This suggests no evidence of residual

MAIC of Sunitinib vs. Everolimus in panNET

confounding or study effects in PFS. It must be noted, however, that this assessment would not identify potential biases that may affect only OS and not PFS.

In conclusion, based on the latest available clinical data from the A6181111 and RADIANT-3 trials, this MAIC showed similar PFS and OS between sunitinib and everolimus in patients with advanced panNETs. Given the differing mechanism of action, patients should have access to both active agents during the course of their treatment, along with awareness of their respective toxicity profiles. There are relatively limited data on the use of and outcomes with these treatments outside of the regulatory setting; a few studies have examined their use and found both to be safe and effective in actual practice.[14-16] Future research is needed to understand the optimal sequencing of these therapies (with respect to each other and other treatment options) as well as identification of biomarkers of benefit (or lack of benefit) and emergence of resistance.

MAIC of Sunitinib vs. Everolimus in panNET

Table 1 Summary of indirect comparison methods considered in this analysis

	Bucher Comparison	Anchored MAIC	Unanchored MAIC
How Does It Work?	Comparison derived from ratio of HRs for sunitinib vs. BSC and everolimus vs. BSC.	Adjusted HR for sunitinib vs. BSC derived after matching population to RADIANT-3. Comparison derived from adjusted HR vs. RADIANT-3 in Bucher analysis.	Comparison derived without use of BSC arms. Compares outcomes (Kaplan–Meier curves) between sunitinib and everolimus, after matching the sunitinib population to RADIANT-3.
Key Assumption	The common BSC arms are comparable between studies. The effect of each treatment is constant across different subgroups (i.e., no effect modification).	All potential effect modifiers are included in matching. Effect derived by matching to RADIANT-3 is applicable to the original sunitinib population.	All predictors and effect modifiers of outcomes are included in the matching. Effect derived by matching to RADIANT-3 is applicable to the original sunitinib population.
Appropriateness for PFS Comparison	Appropriate, assuming no effect modifiers.	Appropriate, and accounts for possible effect modification.	Possible but susceptible to bias due to missing predictors or effect modifiers.
Appropriateness for OS Comparison	BSC arms not comparable due to possibly differential crossover patterns in the two trials.	BSC arms not comparable due to possibly differential crossover patterns in the two trials.	Necessary, to avoid using BSC arms. Susceptible to residual confounding bias.

MAIC of Sunitinib vs. Everolimus in panNET

Table 2 Baseline characteristics in A6181111 before and after matching to RADIANT-3

	A6181111 Unmatched		A6181111 Matched to RADIANT-3			RADIANT-3	
	Sunitinib	BSC+Placebo	Sunitinib	BSC+Placebo Matched to Everolimus*	BSC+Placebo Matched to BSC+Placebo**	Everolimus	BSC+Placebo
N/ESS	86	85	43	31	37	207	203
Age $\geq$ 58 (median in everolimus arm)	0.44	0.47	0.50	0.50	0.50	0.50	0.50
Age $\geq$ 65	0.26	0.27	0.27	0.27	0.27	0.27	0.27
Female	0.51	0.53	0.47	0.47	0.42	0.47	0.42
Race: White	0.56	0.63	0.79	0.79	0.79	0.79	0.79
ECOG PS = 0	0.62	0.48	0.67	0.67	0.66	0.67	0.66
ECOG PS = 1, 2	0.38	0.52	0.33	0.33	0.34	0.33	0.34
Time since diagnosis $\geq$ 3 years	0.40	0.56	0.47	0.47	0.47	0.47	0.47
Number of Sites = 0-1	0.36	0.28	0.25	0.25	0.31	0.25	0.31
Number of Sites = 2	0.36	0.33	0.41	0.41	0.32	0.41	0.32
Number of Sites $\geq$ 3	0.28	0.39	0.34	0.34	0.38	0.34	0.38
Metastasis Site: Liver	0.92	0.92	0.92	0.92	0.92	0.92	0.92
Metastasis Site: Pancreas	0.41	0.36	0.44	0.44	0.41	0.44	0.41
Metastasis Site: Lymph Nodes	0.34	0.48	0.33	0.33	0.36	0.33	0.36
Metastasis Site: Lung	0.10	0.18	0.14	0.14	0.15	0.14	0.15

1 MAIC of Sunitinib vs. Everolimus in panNET

2	3	4	5	6	7	8	9
10	11	12	13	14	15	16	17
18	19	20	21	22	23	24	25
26	27	28	29	30	31	32	33
34	35	36	37	38	39	40	41
42	43	44	45	46	47	48	49
50	51	52	53	54	55	56	57
58	59	60	61	62	63	64	65
66	67	68	69	70	71	72	73
74	75	76	77	78	79	80	81
82	83	84	85	86	87	88	89
90	91	92	93	94	95	96	97
98	99	100	101	102	103	104	105
106	107	108	109	110	111	112	113
114	115	116	117	118	119	120	121
122	123	124	125	126	127	128	129
130	131	132	133	134	135	136	137
138	139	140	141	142	143	144	145
146	147	148	149	150	151	152	153
154	155	156	157	158	159	160	161
162	163	164	165	166	167	168	169
170	171	172	173	174	175	176	177
178	179	180	181	182	183	184	185
186	187	188	189	190	191	192	193
194	195	196	197	198	199	200	201
202	203	204	205	206	207	208	209
210	211	212	213	214	215	216	217
218	219	220	221	222	223	224	225
226	227	228	229	230	231	232	233
234	235	236	237	238	239	240	241
242	243	244	245	246	247	248	249
250	251	252	253	254	255	256	257
258	259	260	261	262	263	264	265
266	267	268	269	270	271	272	273
274	275	276	277	278	279	280	281
282	283	284	285	286	287	288	289
290	291	292	293	294	295	296	297
298	299	300	301	302	303	304	305
306	307	308	309	310	311	312	313
314	315	316	317	318	319	320	321
322	323	324	325	326	327	328	329
330	331	332	333	334	335	336	337
338	339	340	341	342	343	344	345
346	347	348	349	350	351	352	353
354	355	356	357	358	359	360	361
362	363	364	365	366	367	368	369
370	371	372	373	374	375	376	377
378	379	380	381	382	383	384	385
386	387	388	389	390	391	392	393
394	395	396	397	398	399	400	401
402	403	404	405	406	407	408	409
410	411	412	413	414	415	416	417
418	419	420	421	422	423	424	425
426	427	428	429	430	431	432	433
434	435	436	437	438	439	440	441
442	443	444	445	446	447	448	449
450	451	452	453	454	455	456	457
458	459	460	461	462	463	464	465
466	467	468	469	470	471	472	473
474	475	476	477	478	479	480	481
482	483	484	485	486	487	488	489
490	491	492	493	494	495	496	497
498	499	500	501	502	503	504	505
506	507	508	509	510	511	512	513
514	515	516	517	518	519	520	521
522	523	524	525	526	527	528	529
530	531	532	533	534	535	536	537
538	539	540	541	542	543	544	545
546	547	548	549	550	551	552	553
554	555	556	557	558	559	560	561
562	563	564	565	566	567	568	569
570	571	572	573	574	575	576	577
578	579	580	581	582	583	584	585
586	587	588	589	590	591	592	593
594	595	596	597	598	599	600	601
602	603	604	605	606	607	608	609
610	611	612	613	614	615	616	617
618	619	620	621	622	623	624	625
626	627	628	629	630	631	632	633
634	635	636	637	638	639	640	641
642	643	644	645	646	647	648	649
650	651	652	653	654	655	656	657
658	659	660	661	662	663	664	665
666	667	668	669	670	671	672	673
674	675	676	677	678	679	680	681
682	683	684	685	686	687	688	689
690	691	692	693	694	695	696	697
698	699	700	701	702	703	704	705
706	707	708	709	710	711	712	713
714	715	716	717	718	719	720	721
722	723	724	725	726	727	728	729
730	731	732	733	734	735	736	737
738	739	740	741	742	743	744	745
746	747	748	749	750	751	752	753
754	755	756	757	758	759	760	761
762	763	764	765	766	767	768	769
770	771	772	773	774	775	776	777
778	779	780	781	782	783	784	785
786	787	788	789	790	791	792	793
794	795	796	797	798	799	800	801
802	803	804	805	806	807	808	809
810	811	812	813	814	815	816	817
818	819	820	821	822	823	824	825
826	827	828	829	830	831	832	833
834	835	836	837	838	839	840	841
842	843	844	845	846	847	848	849
850	851	852	853	854	855	856	857
858	859	860	861	862	863	864	865
866	867	868	869	870	871	872	873
874	875	876	877	878	879	880	881
882	883	884	885	886	887	888	889
890	891	892	893	894	895	896	897
898	899	900	901	902	903	904	905
906	907	908	909	910	911	912	913
914	915	916	917	918	919	920	921
922	923	924	925	926	927	928	929
930	931	932	933	934	935	936	937
938	939	940	941	942	943	944	945
946	947	948	949	950	951	952	953
954	955	956	957	958	959	960	961
962	963	964	965	966	967	968	969
970	971	972	973	974	975	976	977
978	979	980	981	982	983	984	985
986	987	988	989	990	991	992	993
994	995	996	997	998	999	1000	1001
1002	1003	1004	1005	1006	1007	1008	1009
1010	1011	1012	1013	1014	1015	1016	1017
1018	1019	1020	1021	1022	1023	1024	1025
1026	1027	1028	1029	1030	1031	1032	1033
1034	1035	1036	1037	1038	1039	1040	1041
1042	1043	1044	1045	1046	1047	1048	1049
1050	1051	1052	1053	1054	1055	1056	1057
1058	1059	1060	1061	1062	1063	1064	1065
1066	1067	1068	1069	1070	1071	1072	1073
1074	1075	1076	1077	1078	1079	1080	1081
1082	1083	1084	1085	1086	1087	1088	1089
1090	1091	1092	1093	1094	1095	1096	1097
1098	1099	1100	1101	1102	1103	1104	1105
1106	1107	1108	1109	1110	1111	1112	1113
1114	1115	1116	1117	1118	1119	1120	1121
1122	1123	1124	1125	1126	1127	1128	1129
1130	1131	1132	1133	1134	1135	1136	1137
1138	1139	1140	1141	1142	1143	1144	1145
1146	1147	1148	1149	1150	1151	1152	1153
1154	1155	1156	1157	1158	1159	1160	1161
1162	1163	1164	1165	1166	1167	1168	1169
1170	1171	1172	1173	1174	1175	1176	1177
1178	1179	1180	1181	1182	1183	1184	1185
1186	1187	1188	1189	1190	1191	1192	1193
1194	1195	1196	1197	1198	1199	1200	1201
1202	1203	1204	1205	1206	1207	1208	1209
1210	1211	1212	1213	1214	1215	1216	1217
1218	1219	1220	1221	1222	1223	1224	1225
1226	1227	1228	1229	1230	1231	1232	1233
1234	1235	1236	1237	1238	1239	1240	1241
1242	1243	1244	1245	1246	1247	1248	1249
1250	1251	1252	1253	1254	1255	1256	1257
1258	1259	1260	1261	1262	1263	1264	1265
1266	1267	1268	1269	1270	1271	1272	1273
1274	1275	1276	1277	1278	1279	1280	1281
1282	1283	1284	1285	1286	1287	1288	1289
1290	1291	1292	1293	1294	1295	1296	1297
1298	1299	1300	1301	1302	1303	1304	1305
1306	1307	1308	1309	1310	1311	1312	1313
1314	1315	1316	1317	1318	1319	1320	1321
1322	1323	1324	1325	1326	1327	1328	1329
1330	1331	1332	1333	1334	1335	1336	1337
1338	1339	1340	1341	1342	1343	1344	1345
1346	1347	1348	1349	1350	1351	1352	1353
1354	1355	1356	1357	1358	1359	1360	1361
1362	1363	1364	1365	1366	1367	1368	1369
1370	1371	1372	1373	1374	1375	1376	1377
1378	1379	1380	1381	1382	1383	1384	1385
1386	1387	1388	1389	1390	1391	1392	1393
1394	1395	1396	1397	1398	1399	1400	1401
1402	1403	1404	1405	1406	1407	1408	

MAIC of Sunitinib vs. Everolimus in panNET

Table 3 Summary of PFS results based on reported hazard ratios and the MAIC analyses

	Hazard Ratio	95% Lower Bound	95% Upper Bound
Based on Reported HRs			
Sunitinib vs. BSC + placebo	0.42	0.26	0.66
Everolimus vs. BSC + placebo	0.35	0.27	0.45
Bucher: Sunitinib vs. Everolimus	1.20	0.72	2.01
Anchor-based MAIC			
Sunitinib vs. BSC + placebo—weighted to RADIANT-3	0.30	0.14	0.63
Everolimus vs. BSC + placebo—observed	0.35	0.27	0.45
Bucher: Sunitinib vs. Everolimus	0.85	0.39	1.89

References

1. Klimstra DS, Et Al. In: *WHO Classification of Tumours of the Digestive System* Bosman FT, Carniero F, Hruban RH, Theise N (Ed.^(Eds). IARC Lyon 322-326 (2010).

2. De Wilde RF, Edil BH, Hruban RH, Maitra A. Well-differentiated pancreatic neuroendocrine tumors: from genetics to therapy. *Nature Reviews Gastroenterology and Hepatology* 9 199-208 (2012).

3. Wong MH, Chan DL, Lee A *et al.* Systematic Review and Meta-Analysis on the Role of Chemotherapy in Advanced and Metastatic Neuroendocrine Tumor (NET). *PLoS One* 11(6), e0158140 (2016).

4. Raymond E, Dahan L, Raoul JL *et al.* Sunitinib malate for the treatment of pancreatic neuroendocrine tumors. *N Engl J Med* 364(6), 501-513 (2011).

5. Yao JC, Shah MH, Ito T *et al.* Everolimus for advanced pancreatic neuroendocrine tumors. *N Engl J Med* 364(6), 514-523 (2011).

6. Bucher HC, Guyatt GH, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *J Clin Epidemiol* 50(6), 683-691 (1997).

7. Glenny AM, Altman DG, Song F *et al.* Indirect comparisons of competing interventions. *Health Technol Assess* 9(26), 1-134, iii-iv (2005).

8. Sutton A, Ades AE, Cooper N, Abrams K. Use of indirect and mixed treatment comparisons for technology assessment. *Pharmacoeconomics* 26(9), 753-767 (2008).

9. Signorovitch JE, Wu EQ, Yu AP *et al.* Comparative effectiveness without head-to-head trials: a method for matching-adjusted indirect comparisons applied to psoriasis treatment with adalimumab or etanercept. *Pharmacoeconomics* 28(10), 935-945 (2010).

10. Signorovitch J, Swallow E, Kantor E *et al.* Everolimus and sunitinib for advanced pancreatic neuroendocrine tumors: a matching-adjusted indirect comparison. *Exp Hematol Oncol* 2(1), 32 (2013).

11. Raymond E, Niccoli P, Castellano DE, Et Al. Sunitinib in Patients with Advanced, Progressive Pancreatic Neuroendocrine Tumors: Final Overall Survival Results From a Phase III Randomized Study, Including Adjustment for Crossover. *American Society of Clinical Oncology--Gastrointestinal Cancers Symposium* (2016).

12. Pavel ME, Lombard-Bohas C, Van Cutsem E, Et Al. Everolimus in Patients With Advanced, Progressive Pancreatic Neuroendocrine Tumors: Overall Survival Results From the Phase 3 RADIANT-3 Study After Adjusting for Crossover Bias. Presented at: *American Society of Clinical Oncology Annual Meeting*. Chicago, IL 2015.

13. Hoyle MW, Henley W. Improved curve fits to summary survival data: application to economic evaluation of health technologies. *BMC Med Res Methodol* 11 139 (2011).

14. Rinzivillo M, Fazio N, Pusceddu S *et al.* Sunitinib in patients with pre-treated pancreatic neuroendocrine tumors: A real-world study. *Pancreatology* 18(2), 198-203 (2018).

15. Panzuto F, Rinzivillo M, Fazio N *et al.* Real-world study of everolimus in advanced progressive neuroendocrine tumors. *Oncologist* 19(9), 966-974 (2014).

16. Yoo C, Cho H, Song MJ *et al.* Efficacy and safety of everolimus and sunitinib in patients with gastroenteropancreatic neuroendocrine tumor. *Cancer Chemother Pharmacol* 79(1), 139-146 (2017).

Figure 1. PFS curves for sunitinib and everolimus as observed and after matching on all variables

Figure 2. OS curves for sunitinib and everolimus as observed and after matching on all variables

Figure 3. PFS curves for BSC + placebo groups in the two trials as observed and after matching on all variables

Summary Points

- Phase III trials for sunitinib (A6181111) and everolimus (RADIANT-3) have demonstrated efficacy for the treatment of patients with advanced, progressive, well-differentiated pancreatic neuroendocrine tumors (panNET) when added to best supportive care (BSC); there are no head-to-head studies comparing these treatments.
- A matching-adjusted indirect comparison (MAIC) has been previously performed based on the RADIANT-3 trial and found comparable progression free survival (PFS) and overall survival (OS) with sunitinib and everolimus. More recent data cuts from these trials have become available, allowing an updated comparison of OS.
- Our MAIC was based on the A6181111 trial but otherwise followed a similar methodology to that applied in the previous analysis, matching on all commonly available baseline characteristics of the study populations (including a few that were not considered in the previous MAIC).
- PFS was analyzed with an anchored MAIC using the studies' common BSC arms as an anchor. A6181111 and RADIANT-3 allowed crossover, so an unanchored comparison of OS was deemed more appropriate for this outcome since crossover may have affected OS differentially in the two trials.
- The analyses yielded hazard ratios for sunitinib vs. everolimus of 0.85(0.48-1.55) and 0.82 (0.53–1.27) for PFS and OS, respectively. Confidence intervals of both HRs include the null, indicating comparable outcomes after matching, consistent with findings from the previous MAIC.
- The main limitation of these analyses relates to the smaller size of the A6181111 trial, which after matching on baseline characteristics, leaves a smaller effective sample size and wider confidence intervals from comparative analyses.
- Another limitation on the risk of residual confounding due to factors that were not available for matching, although sensitivity analyses did not yield evidence to suggest this was likely.
- Given the differing mechanism of action and the comparable PFS and OS of sunitinib and everolimus, patients should have access to both active agents during the course of their treatment, along with awareness of their respective toxicity profiles.

Key References

- Raymond E, Dahan L, Raoul JL et al. Sunitinib malate for the treatment of pancreatic neuroendocrine tumors. *N Engl J Med* 364(6), 501-513 (2011).
Original publication of the A6181111 trial findings for sunitinib.
- Yao JC, Shah MH, Ito T et al. Everolimus for advanced pancreatic neuroendocrine tumors. *N Engl J Med* 364(6), 514-523 (2011).
Original publication of the RADIANT-3 trial findings for everolimus.
- Signorovitch J, Swallow E, Kantor E et al. Everolimus and sunitinib for advanced pancreatic neuroendocrine tumors: a matching-adjusted indirect comparison. *Exp Hematol Oncol* 2(1), 32 (2013).
Publication based on the previous MAIC based on the RADIANT-3 trial.
- Raymond E, Niccoli P, Castellano DE, Et Al. Sunitinib in Patients with Advanced, Progressive Pancreatic Neuroendocrine Tumors: Final Overall Survival Results from a Phase III Randomized Study, Including Adjustment for Crossover. *American Society of Clinical Oncology--Gastrointestinal Cancers Symposium* (2016).
Updated data from the A6181111 trial for sunitinib.
- Pavel ME, Lombard-Bohas C, Van Cutsem E, Et Al. Everolimus in Patients with Advanced, Progressive Pancreatic Neuroendocrine Tumors: Overall Survival Results from the Phase 3 RADIANT-3 Study After Adjusting for Crossover Bias. Presented at: *American Society of Clinical Oncology Annual Meeting*. Chicago, IL 2015.
Updated data from the RADIANT-3 trial for everolimus.