Patient-reported outcomes of patients with advanced renal cell carcinoma treated with nivolumab plus ipilimumab versus sunitinib (CheckMate 214): a randomised, phase 3 trial

David Cella, Viktor Grünwald, Bernard Escudier, Hans J Hammers, Saby George, Paul Nathan, Marc-Oliver Grimm, Brian I Rini, Justin Doan, Cristina Ivanescu, Jean Paty, Sabeen Mekan, Robert J Motzer

Summary

Background In the ongoing phase 3, CheckMate 214 trial, nivolumab plus ipilimumab improved overall survival compared with sunitinib in patients with intermediate or poor risk, previously untreated, advanced renal cell carcinoma. We aimed to assess whether health-related quality of life (HRQoL) could be used to further describe the benefit-risk profile of nivolumab plus ipilimumab versus sunitinib.

Methods In the phase 3, randomised, controlled, CheckMate 214 trial, patients aged 18 years and older with previously untreated, advanced or metastatic renal cell carcinoma with a clear-cell component were recruited from 175 hospitals and cancer centres in 28 countries. Patients were categorised by risk status into favourable, intermediate, and poor risk subgroups and randomly assigned (1:1) to open-label nivolumab 3 mg/kg plus ipilimumab 1 mg/kg every 3 weeks for four doses followed by nivolumab 3 mg/kg every 2 weeks, or sunitinib 50 mg/day for 4 weeks of each 6-week cycle. Randomisation was done with a block size of four and stratified by risk status and geographical region. Patient-reported outcomes (PROs) were assessed using the Functional Assessment of Cancer Therapy Kidney Symptom Index-19 (FKSI-19), Functional Assessment of Cancer Therapy-General (FACT-G), and EuroQol five dimensional three level (EQ-5D-3L) instruments. The coprimary endpoints of the trial, reported previously, were overall survival, progression-free survival, and the proportion of patients who had an objective response in those categorised as at intermediate or poor risk. PROs in all randomised participants were assessed as an exploratory endpoint; here we report this exploratory endpoint. This study is registered with ClinicalTrials.gov, number NCT02231749, and is ongoing but is now closed to recruitment.

Findings Between Oct 16, 2014, and Feb 23, 2016, of 1390 patients screened, 1096 (79%) were randomly assigned to treatment, of whom 847 (77%) were at intermediate or poor risk and randomly assigned to nivolumab plus ipilimumab (n=425) or sunitinib (n=422). Median follow-up was 25·2 months (IQR 23·0–27·4). PROs were more favourable with nivolumab plus ipilimumab than sunitinib throughout the first 103 weeks after baseline, with mean change from baseline at week 103 for FKSI-19 total score being 4·00 (95% CI 1·91 to 6·09) for nivolumab plus ipilimumab versus −3·14 (−6·03 to −0·25) for sunitinib (p<0·0001), and for FACT-G total score being 4·77 (1·73 to 7·82) for nivolumab plus ipilimumab versus −4·32 (−8·54 to −0·11) for sunitinib (p=0·0005). Significant differences were also seen for four of five FKSI-19 domains (disease-related symptoms, physical disease-related symptoms, treatment side-effects, and functional wellbeing) and FACT-G physical and functional wellbeing domains. However, there was no significant difference between the treatment groups at week 103 in EQ-5D-3L visual analogue rating scale (VAS) scores, with mean change from baseline to week 103 of 10·07 (95% CI 4·35 to 15·80) for nivolumab plus ipilimumab and 6·40 (−1·36 to 14·16) for sunitinib (p=0·45). Compared with sunitinib, nivolumab plus ipilimumab reduced risk of deterioration in FKSI-19 total score (hazard ratio [HR] 0·54; 95% CI 0·46–0·63), FACT-G total score (0·63, 0·52–0·75), and EQ-5D-3L VAS score (HR 0·75, 95% CI 0·63–0·89) and UK utility scores (0·67, 0·57–0·80).

Interpretation Nivolumab plus ipilimumab leads to fewer symptoms and better HRQoL than sunitinib in patients at intermediate or poor risk with advanced renal cell carcinoma. These results suggest that the superior efficacy of nivolumab plus ipilimumab over sunitinib comes with the additional benefit of improved HRQoL.

Funding Bristol-Myers Squibb and ONO Pharmaceutical.

Copyright 2019 Elsevier Ltd. All rights reserved.

Lancet Oncol 2019; 20: 297–310

Published Online
January 15, 2019
http://dx.doi.org/10.1016/S1470-2045(18)30778-2

See Comment page 179

Northwestern University, Chicago, IL, USA (Prof D Cella PhD); **Hannover Medical School, Hannover, Lower Saxony, Germany** (Prof V Grünwald MD); **Institut Gustave Roussy, Villejuif, France** (Prof B Escudier MD); **UT Southwestern, Dallas, TX, USA** (H J Hammers MD); **Roswell Park Cancer Institute, Buffalo, NY, USA** (S George MD); **Mount Vernon Cancer Centre, Northwood, London, UK** (P Nathan MD); **Department of Urology, Jena University Hospital, Jena, Germany** (Prof M-O Grimm MD); **Cleveland Clinic Taussig Cancer Center, Cleveland, OH, USA** (Prof B I Rini MD); **Bristol-Myers Squibb, Princeton, NJ, USA** (J Doan MPH, S Mekan MD); **IQVIA, Durham, NC, USA** (C Ivanescu PhD, J Paty PhD); **and Memorial Sloan Kettering Cancer Center, New York, NY, USA** (Prof R J Motzer MD)

Correspondence to:
Prof David Cella, Department of Medical Social Sciences, Northwestern University, Chicago, IL 60208, USA
d-cella@northwestern.edu

Introduction

About 30% of patients with renal cell carcinoma have advanced disease, and prognosis is dependent on the presence of established risk factors.[1–4] A validated model designed by the International Metastatic Renal Cell Carcinoma Database Consortium (IMDC) is often used

Research in context

Evidence before this study

We searched PubMed for publications, with no restrictions on article type or language, from database inception until July 20, 2018, using the search terms "nivolumab", "nivolumab AND ipilimumab", "renal cell carcinoma", "RCC", "kidney cancer", "advanced and metastatic RCC", "advanced and metastatic renal cell carcinoma", "health-related quality of life", "HRQoL", "overall survival", "FKSI-DRS", "FACT-G", and "EQ-5D", with specific attention to randomised phase 3 trials with any of the following comparators: cabozantinib, mTOR inhibitors (everolimus, temsirolimus), VEGF inhibitors (sunitinib, sorafenib, bevacizumab, axitinib, pazopanib), or immuno-oncology therapeutics. The only randomised, open-label, phase 3 trial we found in first-line therapy was CheckMate 214, which compared nivolumab plus ipilimumab with sunitinib in patients with previously untreated, advanced or metastatic renal cell carcinoma. We manually searched recent clinical practice guidelines and conference presentations in the field and identified several studies that investigated other therapies for first-line treatment of advanced or metastatic renal cell carcinoma. The phase 3 COMPARZ trial established sunitinib and pazopanib as standards of care for advanced renal cell carcinoma regardless of patient risk group. Other options for first-line therapy in patients at poor risk are tivozanib (in Europe but not the USA) and temsirolimus, but they are used less often than pazopanib and sunitinib in clinical practice. With the exception of temsirolimus, health-related quality of life (HRQoL) data were collected for each of these drugs in their respective phase 3 trials. Finally, cabozantinib is approved by the US Food and Drug Administration (FDA) for first-line treatment of advanced renal cell carcinoma regardless of risk category and by the European Medicines Agency for first-line treatment of patients at intermediate or poor risk based on the results of the phase 2 CABOSUN trial, which did not collect HRQoL data.

Added value of this study

The efficacy results from CheckMate 214 showed that the risk of death was significantly lower among patients with intermediate or poor risk advanced renal cell carcinoma with nivolumab plus ipilimumab than with sunitinib, and the proportion of patients with an objective response was significantly higher with nivolumab plus ipilimumab than with sunitinib. Progression-free survival among these patients was longer with nivolumab plus ipilimumab than with sunitinib, but did not meet the pre-specified boundary for statistical significance. Patient-reported outcomes (PROs) collected using the unctional Assessment of Cancer Therapy Kidney Symptom Index-19 (FKSI-19), Functional Assessment of Cancer Therapy-General (FACT-G), and EuroQol five dimensional three level (EQ-5D-3L) instruments as exploratory endpoints show that nivolumab plus ipilimumab maintained high HRQoL and low symptom burden throughout the trial. Regardless of the PRO instrument and analysis, all PRO-related findings consistently favoured nivolumab plus ipilimumab over sunitinib. HRQoL data contribute to the overall benefit-risk profile of a therapy. In this case, despite more patients discontinuing treatment with nivolumab plus ipilimumab than with sunitinib because of drug-related adverse events, nivolumab plus ipilimumab had a more favourable effect on HRQoL than sunitinib did, suggesting a positive overall risk-benefit profile for the combination treatment.

Implications of all the available evidence

PROs from the CheckMate 214 trial imply that nivolumab plus ipilmumab is associated with improved HRQoL in patients at intermediate or poor risk with advanced renal cell carcinoma compared with sunitinib. On the basis of the CheckMate 214 primary efficacy and safety results, in April, 2018, the US FDA approved the combination of nivolumab plus ipilimumab as first-line therapy for treatment-naive patients at intermediate or poor risk with advanced renal cell carcinoma, establishing a new standard of care in this setting. In March, 2018, the European Association of Urology updated their guidelines to recommend the use of nivolumab plus ipilimumab as first-line therapy in this patient population, and the combination has also been added as a category 1 preferred recommendation for patients at intermediate and poor risk as first-line treatment in the 2018 version 4 National Comprehensive Cancer Network guidelines for kidney cancer. Furthermore, on Jan 11, 2019, the European Medicines Agency approved nivolumab in combination with ipilimumab for first-line treatment of patients with intermediate or poor risk advanced renal cell carcinoma.

to categorise patients as being at favourable, intermediate, or poor prognostic risk on the basis of the following risk factors: Karnofsky performance status, time from diagnosis to first-line systemic therapy, haemoglobin concentrations, neutrophil and platelet counts, and corrected serum calcium concentration.[3,4] Although targeted treatments available in the first-line setting for advanced renal cell carcinoma including sunitinib, pazopanib, bevacizumab plus interferon-α, and temsirolimus (for patients at poor risk only) have improved efficacy and safety compared with previous therapies, these treatments might not or might only modestly improve patients' health-related quality of life (HRQoL).[5–7] Therefore, effective treatment with improved HRQoL for patients with advanced renal cell carcinoma is an unmet medical need in this setting.

Nivolumab is a fully human IgG4 programmed cell death protein 1 (PD-1) immune checkpoint inhibitor antibody approved as monotherapy for treatment of advanced or metastatic renal cell carcinoma after treatment with antiangiogenic therapy.[8] Ipilimumab is a human cytotoxic T-lymphocyte antigen-4 (CTLA-4) immune checkpoint inhibitor antibody that is approved for treatment of metastatic melanoma alone[9] and in

combination with nivolumab.[8,10] After approval of nivolumab as monotherapy, the combination of nivolumab plus ipilimumab gained regulatory approval in the USA by the US Food and Drug Administration (FDA) and other countries as first-line therapy for patients with intermediate or poor risk renal cell carcinoma.[8] Ipilimumab is not approved as monotherapy in the renal cell carcinoma setting.[9]

To our knowledge, CheckMate 214 was the first phase 3 trial to show overall survival superiority of nivolumab plus ipilimumab compared with standard-of-care sunitinib as first-line treatment for patients at intermediate or poor risk with advanced renal cell carcinoma (median overall survival not reached [95% CI 28·2 months–not estimable] *vs* 26·0 months [22·1–not estimable]; hazard ratio [HR] 0·63, 99·8% CI 0·44–0·89; p<0·001).[11] In the intention-to-treat population, which included patients at favourable, intermediate, and poor risk, overall survival was also significantly improved with nivolumab plus ipilimumab compared with sunitinib (median not reached [95% CI not estimable–not estimable] *vs* 32·9 months [95% CI not estimable–not estimable]; HR 0·68, 99·8% CI 0·49–0·95; p=0·0003). A higher incidence of treatment discontinuation due to drug-related adverse events was observed in the nivolumab plus ipilimumab group than in the sunitinib group, raising some concerns regarding the safety and tolerability of nivolumab plus ipilimumab despite the demonstrated superior efficacy. Despite these dropouts, the safety of the combination was manageable in this trial, and substantial differences in patient-reported outcome (PRO) scores by use of the National Comprehensive Cancer Network Functional Assessment of Cancer Therapy–Kidney Symptom Index (FKSI-19) favouring nivolumab plus ipilimumab over sunitinib have been previously reported using descriptive statistics (p<0·001).[11] Here, we describe further PRO results using the FKSI-19, Functional Assessment of Cancer Therapy-General (FACT-G), and EuroQol five dimensional three level (EQ-5D-3L) instruments for the participants in the CheckMate 214 population who were randomly assigned treatment or who were randomly assigned, at intermediate or poor risk, and were assessed for the coprimary endpoints.

Methods

Study design and participants

In this randomised, open-label, phase 3 trial, patients with previously untreated advanced renal cell carcinoma were recruited from 175 hospitals and cancer centres in 28 countries (full list of study sites and countries in the appendix pp 3–5). Full details of the study design have been previously reported.[11] Eligible patients were aged 18 years or older with histological confirmation of advanced or metastatic renal cell carcinoma with a clear-cell component with no previous systemic therapy for their disease, with the exception of one previous adjuvant or neoadjuvant therapy (not including VEGF-targeted agents) for completely resectable renal cell carcinoma if recurrence occurred 6 months or more after the last dose. Further inclusion criteria were a Karnofsky performance status of 70% or more, and measurable disease per Response Evaluation Criteria in Solid Tumors (RECIST) version 1.1. Key exclusion criteria were history of or current central nervous system metastases and active or recent history of autoimmune disease or use of systemic corticosteroids within 14 days of group assignment. Full inclusion and exclusion criteria have been previously reported.[11] Eligible patients were categorised by risk status according to IMDC prognostic factors into favourable (0 risk factors), intermediate (1–2 risk factors), and poor (3–6 risk factors) risk subgroups at randomisation according to the number of the following risk factors present: Karnofsky performance status score of 70%, time from initial diagnosis to randomisation less than 1 year, haemoglobin concentration below the lower limit of the normal range, corrected serum calcium concentration of more than 10 mg/dL (2·5 mmol/L), absolute neutrophil count above the upper limit of normal (ULN), and platelet count above the ULN; upper and lower limits of normal were defined according to local laboratory ranges.[4,11] By IMDC risk category, patients at intermediate risk who received standard of care have been reported to have a median overall survival of 22·5 months (95% CI 18·7–25·1) and those at poor risk a median overall survival of 7·8 months (95% CI 6·5–9·7).[3] The primary efficacy population in CheckMate 214 was patients with intermediate or poor risk.

All patients provided written informed consent on the basis of the Declaration of Helsinki principles. This trial was approved by the institutional review board at each site and conducted according to Good Clinical Practice guidelines defined by the International Conference on Harmonisation. The protocol is available in the appendix.

Randomisation and masking

Participants were randomly assigned 1:1 by use of an interactive voice-response system to receive either nivolumab plus ipilimumab or sunitinib. Randomisation was completed with a block size of four and was stratified by IMDC prognostic score (0 *vs* 1–2 *vs* 3–6) and geographical region (USA *vs* Canada and Europe *vs* the rest of the world). The trial was open label; thus patients and investigators were not masked to treatment assignment.

Procedures

See Online for appendix

The treatment regimen was administered in cycles of 6 weeks. Participants assigned to nivolumab plus ipilimumab were given nivolumab 3 mg/kg over 60 min intravenously and ipilimumab 1 mg/kg over 30 min intravenously once every 3 weeks for four doses over

two cycles (induction phase), followed by nivolumab monotherapy 3 mg/kg every 2 weeks for the remainder of treatment (maintenance phase). Participants assigned to sunitinib were given sunitinib at 50 mg orally once daily for 4 weeks on and 2 weeks off.[11] Treatment was continued until disease progression or unacceptable toxicity. As per a protocol amendment on Nov 13, 2017, patients could be treated indefinitely, choose to switch to a flat dose of nivolumab 240 mg every 2 weeks if they were receiving nivolumab 3 mg/kg every 2 weeks maintenance, or choose to discontinue after 2 years in the absence of disease progression or unacceptable toxicity. No dose reductions were allowed in the nivolumab plus ipilimumab group, whereas two dose reductions were permitted for sunitinib in 12·5 mg increments for grade 3–4 toxic side-effects.[11] Patients could continue study treatment after initial RECIST-defined progression as assessed by an investigator if the investigator thought the patient tolerated and benefited from the treatment despite initial evidence of disease progression. Disease assessments were done with CT or MRI at baseline, 12 weeks after randomisation, and every 6 weeks for the first 13 months, then every 12 weeks until progression or treatment discontinuation. Adverse events were assessed continuously during treatment visits, then at follow-up visits 1 and 2 and during overall survival follow-up, and graded according to the National Cancer Institute Common Terminology Criteria for Adverse Events (CTCAE) version 4.0 and reported from the first dose and up to and including 100 days after the last dose of study treatment. Laboratory assessments of blood chemistry were done on day 1 of week 1 and day 1 of week 4 during cycles 1 and 2; day 1 of week 1, day 1 of week 3, and day 1 of week 5 during subsequent cycles; and at follow-up visits 1 and 2.

PROs were assessed by use of the FKSI-19,[12] FACT-G,[13] and EQ-5D-3L[14] instruments (appendix p 7). The FKSI-19 is a validated 19-item instrument that measures tumour-specific PROs in patients with kidney cancer. Patients rate their symptoms on a five-point scale, with responses ranging from "not at all" to "very much". The FKSI-19 contains the following domains: disease-related symptoms (DRS), disease-related symptoms physical (DRS-P), disease-related symptoms emotional (DRS-E), treatment side-effects (TSE), and functional wellbeing (FWB). The FKSI-19 total score is based on all 19 items and ranges from 0–76, with higher scores indicating fewer symptoms.[12,15]

The FACT-G is a validated 27-item instrument that measures general cancer HRQoL and is divided into four primary domains: physical wellbeing (PWB), social and family wellbeing (SWB), emotional wellbeing (EWB), and FWB. The FACT-G total score is based on all 27 items and ranges from 0–108, with higher scores indicating better HRQoL.[13]

The EQ-5D-3L is a validated, standardised instrument for measuring general health status that includes the following five domains: mobility, self-care, usual activities, pain and discomfort, and depression and anxiety, and has a visual analogue rating scale (VAS). The VAS uses a 100-point scale, in which patients self-rate their health state with zero being the worst health imaginable and 100 the best health imaginable.[14] An EQ-5D-3L health-state utility index score ranging from 0 to 1 is calculated for each of the health states described by the instrument on the basis of values provided in large general population studies. In this trial, the UK preference weights were applied.[16]

All PRO instruments were administered on day 1 of week 1 and day 1 of week 4 of the first two cycles; day 1 of week 1 and day 1 of week 5 of the next two cycles; day 1 of week 1 of the subsequent cycles; and at the first two follow-up visits. Follow-up visit one was 30 days (or within 7 days before or after this timepoint) after the last dose or on the date of discontinuation (or within 7 days before or after this timepoint) if discontinuation occurred more than 37 days after the last dose. Follow-up visit two was 84 days (or within 7 days before or after this timepoint) after follow-up visit one. During the follow-up period, administration of EQ-5D-3L was additionally scheduled every 3 months for the first 12 months and every 6 months thereafter (figure 1). The PRO assessment schedule was designed such that PROs were collected at each visit that ipilimumab was given to capture the effects of combination treatment. PROs were collected less often during the nivolumab monotherapy maintenance phase to avoid cognitive exhaustion in patients from overly frequent assessments. All questionnaires were paper based and were administered in person by trial site staff before study treatment. Questionnaires were developed assuming an at least primary-school reading level. For follow-up visits, sites were permitted to administer questionnaires via telephonic scripts if necessary, although in person was still preferred.

Outcomes

The coprimary endpoints of CheckMate 214, reported previously,[11] were overall survival, progression-free survival, and the proportion of patients who had an objective response per independent central radiology review committee in the primary efficacy population of patients with intermediate or poor risk (according to IMDC definitions) previously untreated advanced or metastatic renal cell carcinoma.[11] Here we report the exploratory endpoints of PROs, which included evaluation of disease-related symptoms based on the FKSI-19 scale, HRQoL using the FACT-G instrument, and assessment of changes in global health status using the EQ-5D-3L instrument. We had planned to compare differences in PROs between groups at 6 months; however, post-hoc we decided to expand our analysis to 103 weeks to analyse the full follow-up data.

Statistical analysis

A detailed description of the statistical analyses for the primary and secondary endpoints has been previously

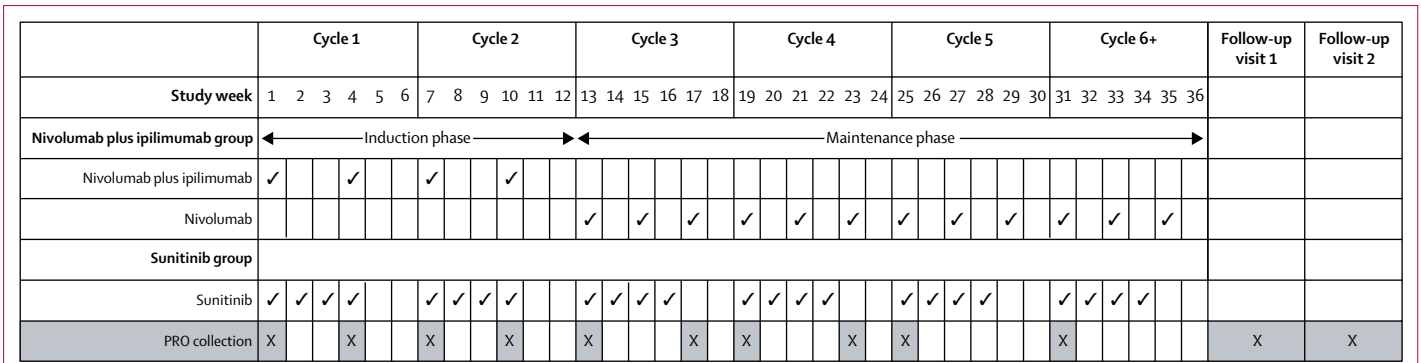

Figure 1: **Schedule of treatment and PRO assessments**

Participants assigned to nivolumab plus ipilimumab were given nivolumab and ipilimumab once every 3 weeks for four doses over two cycles (induction phase), followed by nivolumab monotherapy for the remainder of treatment (maintenance phase). Participants assigned to sunitinib were given sunitinib once daily for 4 weeks on and 2 weeks off. Treatment was continued until disease progression or unacceptable toxicity. Follow-up visit one was 30 days (or within 7 days before or after this timepoint) from the last dose, or on the date of discontinuation (or within 7 days before or after this timepoint) if discontinuation was more than 37 days after the last dose. Follow-up visit two was 84 days (or within 7 days before or after this timepoint) after follow-up visit one. During follow-up, administration of EQ-5D-3L was additionally scheduled every 3 months for the first 12 months and every 6 months thereafter. PRO=patient-reported outcome. EQ-5D-3L=EuroQol five dimensional three level.

reported.[11] Briefly, we estimated that 1070 participants would be required to enrol 820 who were at intermediate or poor risk—ie, the number needed for robust statistical analysis. The overall α level of 0·05 was split among the coprimary endpoints (0·001 for the proportion of patients who achieved an objective response, 0·009 with a power of 80% for progression-free survival, and 0·04 with 90% power for overall survival). PROs were assessed in all participants who were randomly assigned to treatment. Here, we report PRO analyses with a focus on all participants who were randomly assigned to treatment groups and were at intermediate or poor risk, consistent with the primary efficacy population and the FDA-approved indication for nivolumab plus ipilimumab. We also report here the PRO results for the intention-to-treat population—ie, all patients who were randomly assigned to nivolumab plus ipilimumab or sunitinib and who had favourable, intermediate, or poor risk. The proportion of participants who completed the assessment at every timepoint for all PRO instruments was defined as the number and proportion of participants who completed evaluable forms among those who were expected to have PRO assessments—ie, who were alive and still on study. For FKSI-19 and FACT-G, a form was defined as evaluable if at least 80% of the items were completed according to the scoring algorithms for the instruments. For EQ-5D-3L, a form was defined as evaluable if all items were completed. For all analyses, p values less than 0·05 were judged to be nominally significant without multiplicity adjustment.

Our PRO analyses included descriptive statistics for observed scores and change from baseline for each treatment group and timepoint. Although the descriptive analyses for FKSI-19 total score have been previously reported,[11] those analyses used a 4-week window and data were provided for every 4-week period regardless of when the clinic visit took place.[11] By contrast, here we used an analysis window based on the protocol-defined schedule of study visits; therefore, our reported findings differ slightly from those reported previously.

We defined clinically meaningful deterioration of PROs using prespecified threshold values on the basis of score changes from baseline that patients had considered clinically meaningful in previous literature[17–19] or, if no thresholds were available, we calculated the threshold to be half the SD of the baseline values for all participants (appendix p 7).

We defined time to first deterioration as the time from the date of randomisation to the date of the first clinically meaningful deterioration in PRO scores of at least one threshold unit (appendix p 7) compared with the baseline score. Participants who did not have a first deterioration were censored at the date of the last instrument assessment (ie, date of the last non-missing value). Death was not included in the definition of first deterioration; therefore, participants who died and did not have a first deterioration before death were censored at the last completed assessment. Participants with no baseline assessment were censored at the date of randomisation.

To estimate longitudinal changes in PRO scores from baseline at each scheduled visit while on study drug, we used a mixed model for repeated measures (MMRM) analysis, which is one specific analysis of a large group of likelihood-based mixed-effects analyses.[20,21] MMRM is frequently used as the primary PRO analysis method and has been used previously to analyse PRO data in renal cell carcinoma and other indications.[22–24] The dependent variable was change in PRO score from baseline, and the fixed effects were treatment, study visit, randomisation factors (IMDC prognostic score and region) as categorical parameters, baseline PRO score as a continuous parameter, and the interactions between visit and treatment

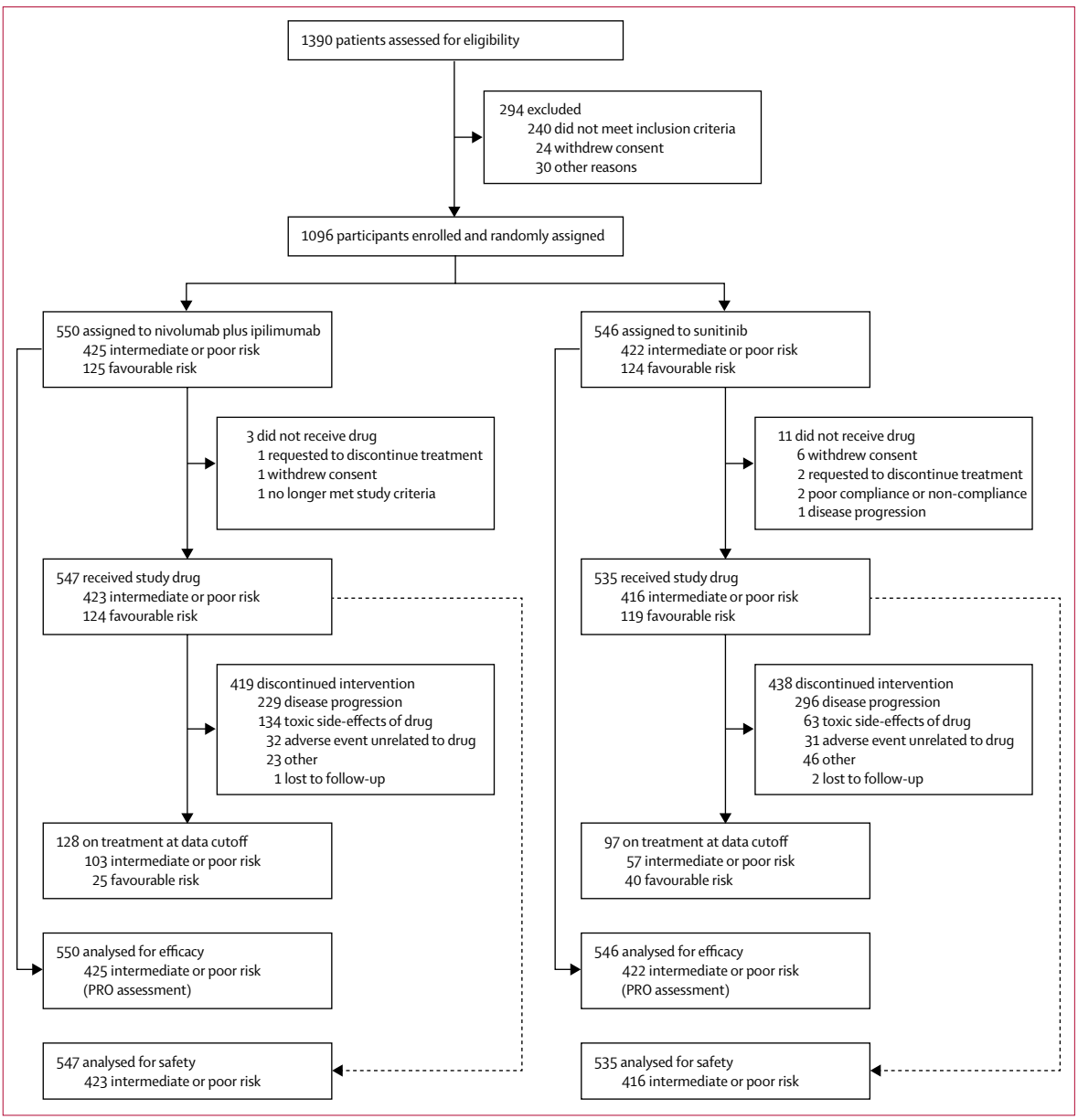

Figure 2: **Trial profile**
PRO=patient-reported outcome.

and between baseline PRO score and visit. We used an unstructured variance–covariance matrix to model the covariance structure among each participant's repeated measures. The prespecified MMRM analysis was limited to the first 103 weeks (timeframe determined post-hoc) after baseline because of the small sample size in both groups beyond this point. MMRM analyses only included on-treatment assessments and assumed that the missing observations (including deaths) were missing at random. To account for missing PRO data based on different reasons for discontinuation of study drug, which therefore might not be missing at random, we did a sensitivity analysis with a pattern mixture model (PMM) involving sequential modelling with multiple imputation by reasons for discontinuation before 103 weeks (appendix p 6).[20] Plots of change in PRO score from baseline by treatment and study visit for the MMRM and PMM analyses report the least-squares mean and standard error.

We included all PRO assessments (including the follow-up visits after treatment discontinuation) in our time-to-event analyses. We used the Kaplan-Meier method to estimate the distribution of time to first deterioration in PRO score. We estimated the HR and

associated 95% CIs using a stratified Cox regression model with treatment as the only covariate and the randomisation factors (IMDC score and geographical region) as strata. Additionally, we did a sensitivity analysis for time to first deterioration using additional threshold values (appendix p 7).

We did prespecified analyses of associations between longitudinal changes in PROs and overall survival and progression-free survival with multivariate time-dependent Cox regression analysis for longitudinal and time-to-event data. In addition to the time-varying covariate (the PRO score), the models included effects for the treatment group and randomisation factors. We fitted separate models for each PRO instrument. We calculated the key measure of the HR and associated 95% CIs for the PRO variable as the hazard of progression-free survival or overall survival per clinically meaningful score change (appendix p 7). These minimally important differences were defined as follows: a one-point increment in FKSI-19 disease-related symptoms-emotional, treatment side-effects, and functional wellbeing scores; a three-points increment in the FKSI-19 total score and disease-related symptoms, and FACT-G physical wellbeing, functional wellbeing, emotional wellbeing, and family or social wellbeing scores; a four-points increment in FKSI-19 disease-related symptoms-physical score; and a seven-points increment in FACT-G total score.

Post-hoc, we used the Kaplan-Meier method to estimate the median overall survival for four groups of participants (irrespective of treatment group) categorised on the basis of FKSI-19 scores at baseline and up to and including week 25, defined as follows: high baseline score, improved; high baseline score, not improved; low baseline score, improved; and low baseline score, not improved. A high baseline score was defined as at or above the median baseline score, and a low baseline score was defined as below the median baseline score. We defined an improvement in HRQoL as a clinically important change of at least 3 points from baseline in FKSI-19 total score. Participants in the not improved groups did not show the change of 3 points or more from baseline in FKS-19 total score. These groups therefore represent patients who showed either maintenance or deterioration in HRQoL. We restricted this analysis to changes from baseline up to and including week 25 to ensure that a large proportion of patients informed the analysis and only patients with a baseline score and at least one score after baseline were included. At week 25, approximately 50% of the total patient population (ie, the population who started treatment at week 0) in both treatment groups had available PRO data. We also did a post-hoc analysis of change in PRO scores for FKSI-19 and FACT-G total scores from baseline for patients with high (ie, those with a baseline score at or above median) and low (ie, those with baseline score below median) PRO scores at baseline to examine how PRO scores changed over time depending on baseline PRO score. All analyses discussed hereon were completed using data from all participants who were randomly assigned to treatment with evaluable questionnaires.

We analysed data using SAS version 9.4. CheckMate 214 is registered with ClinicalTrials.gov, number NCT02231749.

	Intention-to-treat population		Patients at intermediate or poor risk	
	Nivolumab plus ipilimumab (n=550)	Sunitinib (n=546)	Nivolumab plus ipilimumab (n=425)	Sunitinib (n=422)
Age, years				
Median	62 (55–68)	62 (55–68)	62 (54–68)	61 (54–67)
Sex				
Male	413 (75%)	395 (72%)	314 (74%)	301 (71%)
Female	137 (25%)	151 (28%)	111 (26%)	121 (29%)
IMDC risk group				
Favourable	125 (23%)	124 (23%)	0	0
Intermediate	334 (61%)	333 (61%)	334 (79%)	333 (79%)
Poor	91 (17%)	89 (16%)	91 (21%)	89 (21%)
Geographical region				
USA	154 (28%)	153 (28%)	112 (26%)	111 (26%)
Europe and Canada	201 (37%)	199 (36%)	148 (35%)	146 (35%)
Rest of the world	195 (35%)	194 (36%)	165 (39%)	165 (39%)
Quantifiable tumour PD-L1 expression*				
<1%	386/499 (77%)	376/503 (75%)	284/384 (74%)	278/392 (71%)
≥1%	113/499 (23%)	127/503 (25%)	100/384 (26%)	114/392 (29%)
Most common sites of metastasis				
Lung	381 (69%)	373 (68%)	294 (69%)	296 (70%)
Lymph node	246 (45%)	268 (49%)	190 (45%)	216 (51%)
Bone†	112 (20%)	119 (22%)	95 (22%)	97 (23%)
Liver	99 (18%)	107 (20%)	88 (21%)	89 (21%)
Previous nephrectomy				
Yes	453 (82%)	437 (80%)	341 (80%)	319 (76%)
No	97 (18%)	109 (20%)	84 (20%)	103 (24%)
Number of sites with target or non-target lesions‡				
1	123 (22%)	118 (22%)	90 (21%)	84 (20%)
≥2	427 (78%)	427 (78%)	335 (79%)	337 (80%)

Data are median (IQR), n (%), or n/N (%). *Quoted as number divided by total number with evaluable data. †Includes patients who had bone metastases with or without a soft-tissue component. ‡Number of target or non-target lesions at baseline were not reported for one patient in the sunitinib group. IMDC=International Metastatic Renal Cell Carcinoma Database Consortium. PD-L1=programmed cell death ligand-1.

Table 1: **Demographic and baseline characteristics**

Role of the funding source

The funders contributed to the study design, data analysis, and data interpretation in collaboration with the authors. The funders had no role in data collection. Financial support for editorial and writing assistance was provided by the funders. All authors had full access to the data in the study and the corresponding author

	Score ranges	Nivolumab plus ipilimumab group (n=425)	Sunitinib group (n=422)
FSKI-19			
Total score	0–76	60·07 (9·81)	59·05 (10·43)
DRS	0–36	30·74 (4·46)	30·13 (5·12)
DRS-P	0–48	38·97 (6·28)	38·30 (6·91)
DRS-E	0–4	2·17 (1·26)	2·05 (1·34)
TSE	0–12	11·22 (1·34)	11·21 (1·34)
FWB	0–12	7·75 (3·35)	7·49 (3·43)
FACT-G			
Total score	0–108	82·58 (15·04)	80·46 (15·79)
PWB	0–28	23·48 (4·71)	23·35 (4·87)
FWB	0–28	18·75 (6·44)	18·14 (6·66)
EWB	0–24	17·56 (4·37)	16·72 (4·71)
SWB	0–28	22·78 (5·27)	22·29 (5·28)
EQ-5D-3L			
VAS	0–100	70·45 (25·14)	69·61 (26·71)
Utility index	0–1	0·77 (0·25)	0·78 (0·25)

Data are mean (SD). Higher scores indicate better HRQoL (for FACT-G and FKSI-19), fewer symptoms (for FKSI-19), and better health status (EQ-5D-3L). EQ-5D-3L=EuroQoL five dimension three level. HRQoL=health-related quality of life. FKSI-19=functional assessment of cancer therapy-kidney symptom index. DRS=disease-related symptoms. DRS-P=disease-related symptoms physical. DRS-E=disease-related symptoms emotional. TSE=treatment side-effects. FWB=functional wellbeing. FACT-G=functional assessment of cancer therapy-general. PWB=physical wellbeing. EWB=emotional wellbeing. SWB=social and family wellbeing. VAS=visual analogue rating scale.

Table 2: **Baseline HRQoL cores for all randomised participants at intermediate or poor risk**

had final responsibility for the decision to submit for publication.

Results

Between Oct 16, 2014, and Feb 23, 2016, 1390 patients were assessed for eligibility and 1096 (79%) were randomly assigned to nivolumab plus ipilimumab (n=550) or sunitinib (n=546) in CheckMate 214. Of these participants, 847 (77%) were categorised as at intermediate or poor risk and randomly assigned to nivolumab plus ipilimumab (n=425) or sunitinib (n=422; figure 2). Demographic and baseline characteristics of all participants are reported elsewhere,[11] and participants were well balanced between treatment groups (table 1). At the time of database lock (Aug 7, 2017), the median follow-up was 25·2 months (IQR 23·0–27·4).

The number of evaluable baseline questionnaires completed among the 425 participants in the nivolumab plus ipilimumab group was 413 (97%) FKSI-19 questionnaires, 412 (97%) FACT-G questionnaires, and 415 (98%) EQ-5D-3L questionnaires; and among the 422 participants in the sunitinib group was 400 (95%) FKSI-19 questionnaires, 400 (95%) FACT-G questionnaires, and 403 (96%) EQ-5D-3L questionnaires. The proportion of participants who completed the assessments (among those expected to complete them) was more than 80% for all PRO instruments in both treatment groups throughout all visits, except at week 121 when the proportion was 75% in the sunitinib group (appendix p 8). Baseline PRO scores in patients with intermediate or poor risk were similar between the treatment groups and showed few symptoms (FKSI-19), good HRQoL (FACT-G), and good health status (EQ-5D-3L VAS and utlity index; table 2).

The median duration of treatment in all patients who received a trial drug was 7·9 months (95% CI 6·5–8·4) with nivolumab plus ipilimumab and 7·8 months (95% CI 6·4–8·5) with sunitinib. In the randomised population, 157 (29%) of 550 participants in the nivolumab plus ipilimumab and 129 (24%) of 546 in the sunitinib group were treated beyond initial investigator-assessed RECIST-defined progression, as permitted according to the protocol; therefore, a proportion of patients could have had HRQoL assessments after disease progression. In the nivolumab plus ipilimumab group, participants were treated for a median of 3·88 months (IQR 1·38–10·48; maximum 26·1 months) after progression before discontinuation for second progression or another reason, and in the sunitinib group they were treated a median of 1·22 months (IQR 0·53–3·98; maximum 22·1 months) after progression.

Almost all PRO scores after baseline were higher in participants in the nivolumab plus ipilimumab group than in those in the sunitinib group. Most exceptions were after week 103, when the mean score for specific domains favoured sunitinib (FKSI-19 disease-related symptoms and functional wellbeing, and FACT-G social and family wellbeing and functional wellbeing; appendix p 9).

The results of the MMRM analyses are shown in figure 3. Based on MMRM model data, mean change in FKSI-19 score, FACT-G score, and EQ-5D-3L utility index score was higher with nivolumab plus ipilimumab than with sunitinib at most timepoints assessed, but not for EQ-5D-3L VAS (figure 3). Least-squares mean change in score for all instruments was improved in the nivolumab plus ipilimumab group as early as week 10 and was maintained until week 103. This improvement is shown in the change of 4·00 (95% CI 1·91 to 6·09) from baseline at week 103 for nivolumab plus ipilimumab and −3·14 (−6·03 to −0·25) for sunitinib (p<0·0001; figure 3A) for FKSI-19 total score; 4·77 (1·73 to 7·82) for nivolumab plus ipilimumab and −4·32 (−8·54 to −0·11) for sunitinib (p=0·0005; figure 3B) for FACT-G total score; 10·07 (4·35 to 15·80) for nivolumab plus ipilimumab and 6·40 (−1·36 to 14·16) for sunitinib (p=0·45; figure 3C) for EQ-5D-3L VAS; and 0·04 (−0·01 to 0·09) for nivolumab plus ipilimumab and −0·04 (−0·11 to 0·03) for sunitinib (p=0·07; figure 3D) for EQ-5D-3L UK utility index. Significant differences in domain scores showing improved HRQoL with nivolumab plus ipilimumab

compared with sunitinib were also observed at most visits for FKSI-19 disease-related symptoms, physical disease-related symptoms, treatment side-effects, and functional wellbeing (but not for FKSI-19 emotional disease-related symptoms), and for FACT-G physical wellbeing and functional wellbeing (but not for FACT-G emotional wellbeing or family and social wellbeing; appendix pp 10–11). Changes in EQ-5D-3L UK utility index scores were consistently better in the nivolumab plus ipilimumab group than in the sunitinib group during the first 103 weeks (figure 3D); the difference between treatment groups in EQ-5D-3L VAS was significant at only a few timepoints (p<0·05; figure 3C). We did not observe any significant differences favouring sunitinib for any of the PRO instruments at any time during the trial (figure 3; appendix pp 10–11). Generally similar results were observed in the intention-to-treat population (appendix p 12). Results of the PMM sensitivity analysis in the participants at intermediate or poor risk closely align with the MMRM results (appendix p 13).

Analyses of time to first deterioration among participants at intermediate or poor risk showed that combination treatment with nivolumab plus ipilimumab significantly reduced the risk of deterioration in the FKSI-19 total score and disease-related symptoms, physical disease-related symptoms, treatment side-effects, and functional wellbeing domain scores compared with sunitinib, although there was no significant difference between the groups in time to deterioration in scores for emotional disease-related symptoms (figure 4A). Compared with sunitinib, nivolumab plus ipilimumab combination treatment was also associated with a reduced risk of deterioration in FACT-G total score, physical wellbeing, and functional wellbeing scores, but there was no significant difference in time to deterioration in FACT-G emotional wellbeing or social and family wellbeing scores (figure 4B). Additionally, nivolumab plus ipilimumab significantly decreased the risk of deterioration in EQ-5D-3L VAS scores (median time to deterioration 4·2 months [95% CI 2·9–5·1] in the nivolumab plus ipilimumab group *vs* 2·1 months [1·4–2·3] in the sunitinib group; HR 0·75, 95% CI 0·63–0·89) and UK utility index VAS scores (median

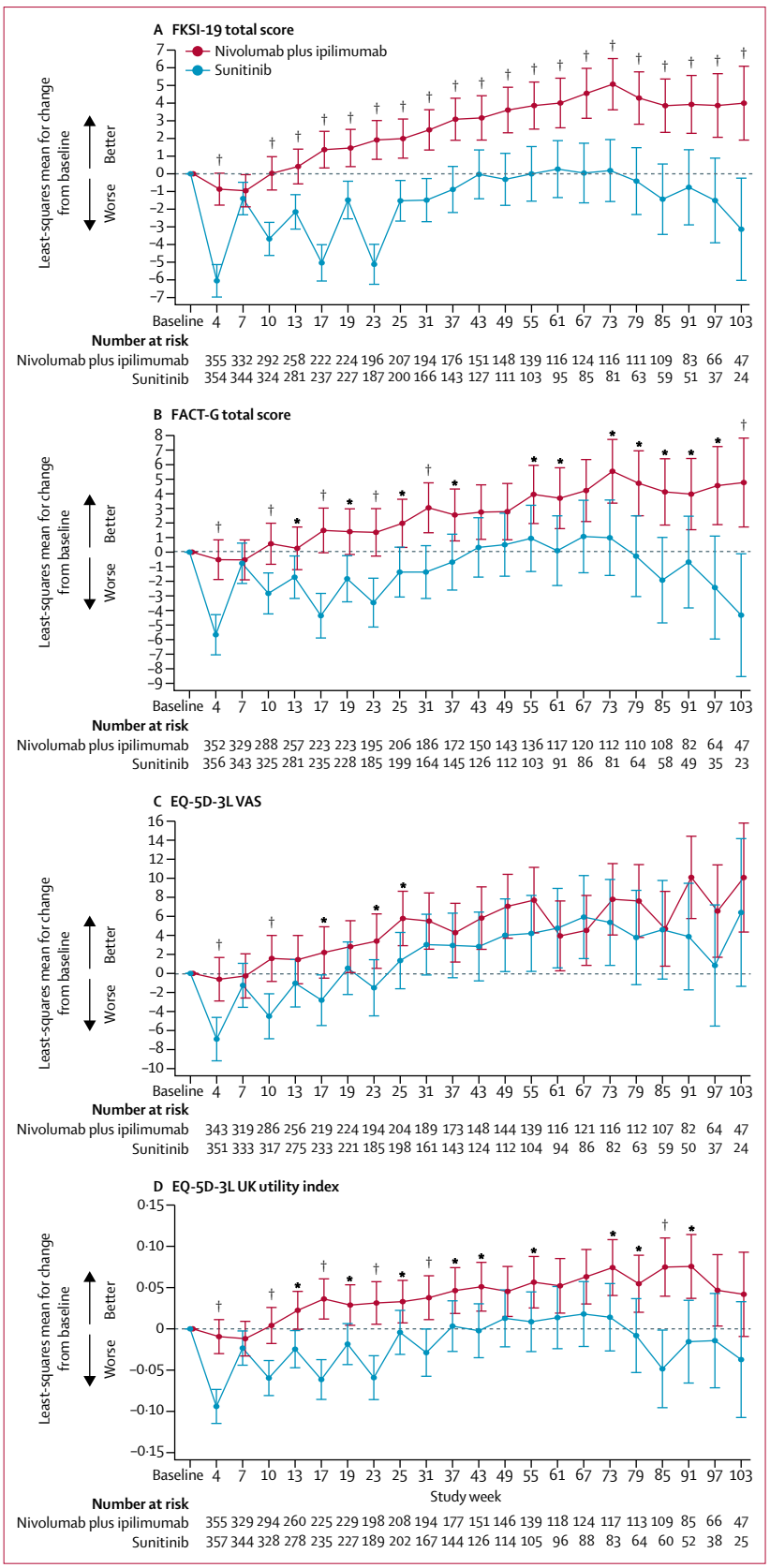

Figure 3: **Mean change from baseline FKSI-19 total score (A), FACT-G total score (B), EQ-5D-3L VAS (C), and EQ-5D-3L UK utility index (D) in participants with intermediate or poor risk**

Data points are least-squares mean change from baseline, and error bars are standard errors and are from mixed-model repeated measures analysis. Mean change from baseline figures for FKSI-19 and FACT-G individual domains are in the appendix (pp 10–11). The number of patients at each timepoint are those with a non-missing score with an evaluable questionnaire. EQ-5D-3L=EuroQoL five dimension three level. FKSI-19=functional assessment of cancer therapy-kidney symptom index-19. FACT-G=functional assessment of cancer therapy-general. VAS=visual analogue scale. *Significant difference between treatment groups (p<0·05). †Significant difference between treatment groups (p<0·001).

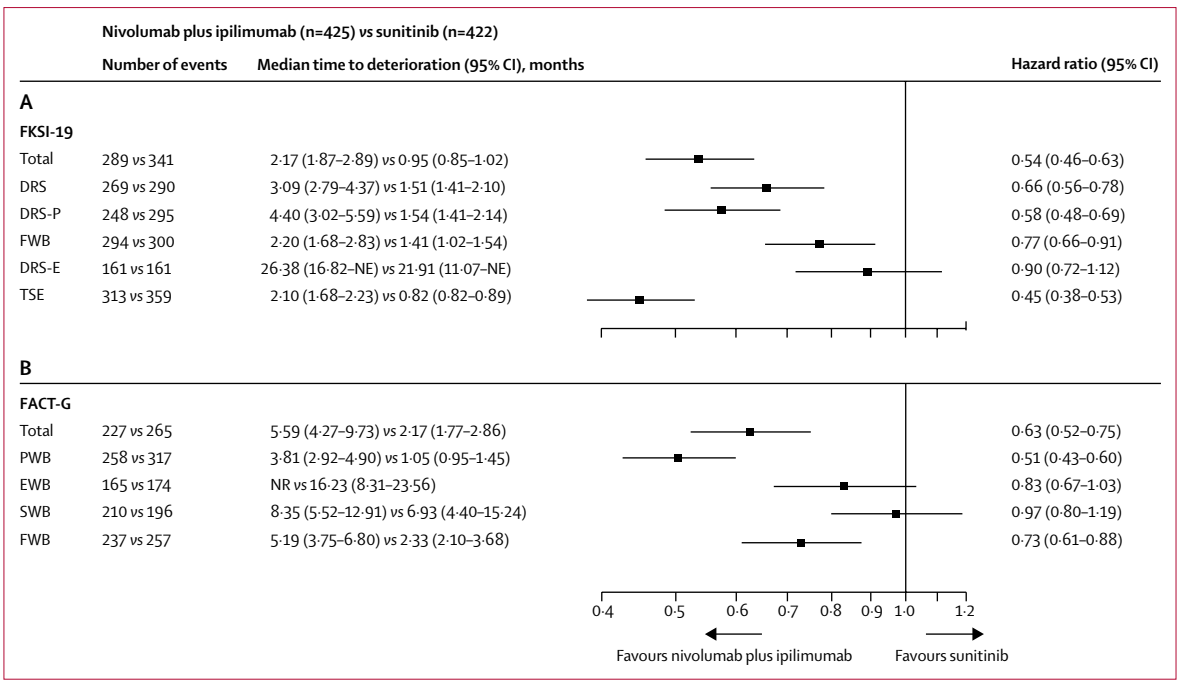

Figure 4: **Time to deterioration in total and domain scores for FKSI-19 (A) and FACT-G (B) instruments among participants with intermediate or poor risk in the CheckMate 214 trial**

Data points are hazard ratios and error bars are 95% CIs. A clinically meaningful deterioration was defined as a decrease of at least one threshold unit compared with baseline; threshold values are in the appendix (p 7). FKSI-19=functional assessment of cancer therapy-kidney symptom index-19. DRS=disease-related symptoms. DRS-P=disease-related symptoms-physical. FWB=functional wellbeing. DRS-E=disease-related symptoms-emotional. NE=not estimable. TSE=treatment side-effects. FACT-G=functional assessment of cancer therapy-general. PWB=physical wellbeing. EWB=emotional wellbeing. NR=not reached. SWB=social and family wellbeing.

time to deterioration 4·9 months [95% CI 3·7–7·0] in the nivolumab plus ipilimumab group *vs* 2·2 months [1·6–2·4] in the sunitinib group; HR 0·67, 0·57–0·80) compared with sunitinib. Our sensitivity analysis defining deterioration of PROs as a change in score equal to or greater than half the SD of the baseline values for all patients yielded similar results to the primary analysis (data not shown). Furthermore, similar time to deterioration results were observed in the intention-to-treat population (appendix p 14).

From the Cox regression analyses, we found that increases in FKSI-19 scores (ie, improvements) were associated with decreased risk of death and disease progression (figure 5A). Similarly, increases in FACT-G total score and physical wellbeing, functional wellbeing, and emotional wellbeing domain scores were associated with reduced risk of death and disease progression (figure 5B). However, increases in FACT-G family and social wellbeing scores were not significantly associated with risk of death or disease progression (figure 5B).

In our post-hoc analysis, the association between high (at or above the median) and low (below the median) FSKI-19 total baseline scores, change in FKSI-19 total scores up to 25 weeks (improved or not improved), and overall survival was examined (appendix p 15). 771 participants with a baseline and at least one score after baseline were included in this analysis. Overall survival was longest in participants with high baseline FKSI-19 scores and FKSI-19 improved from baseline (median overall survival not reached). In participants with high baseline FKSI-19 scores and either FKSI-19 improved or not improved from baseline, the median overall survival was not reached. The shortest median overall survival was in participants with low baseline FKSI-19 scores and FKSI-19 not improved from baseline (16·6 months, 95% CI 12·4–19·9). We also recorded improvements in both FSKI-19 total and FACT-G total scores from baseline with nivolumab plus ipilimumab both in patients with a low baseline score and those with a high baseline score (appendix pp 16–17).

Discussion

Our analysis of PROs in the randomised, open-label, phase 3 CheckMate 214 trial focused on the participants at intermediate or poor risk, consistent with the primary efficacy population and the FDA–approved indication for nivolumab plus ipilimumab. Baseline FACT-G total scores suggested that these participants had good HRQoL, which was further supported by the baseline FKSI-19 functional wellbeing scores. Additionally, these participants had low symptom burden and good health status at baseline. Overall, PRO scores were maintained or improved from baseline with nivolumab plus ipilimumab throughout the trial. Regardless of the PRO instrument or the method of

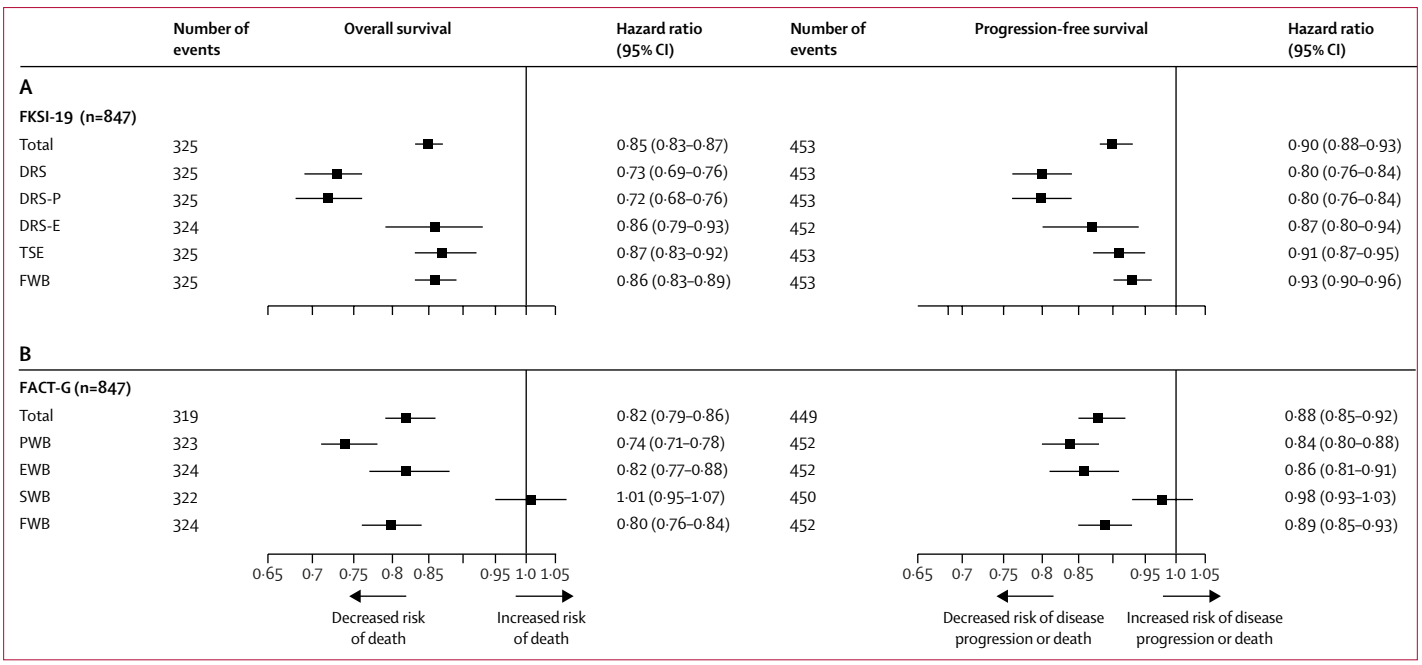

Figure 5: **Association between change in total and domain scores and overall and progression-free survival for FKSI-19 (A) and FACT-G (B) questionnaires among participants with intermediate or poor risk in the CheckMate 214 trial**

Data are from Cox regression analysis. Hazard ratios correspond to a 1-point increment in FKSI-19 DRS-E, TSE, and FWB scores; a 3-point increment in the FKSI-19 total and DRS, and FACT-G PWB, FWB, EWB, and SWB scores; a 4-point increment in FKSI-19 DRS-E score; and a 7-point increment in FACT-G total score (appendix). FKSI-19=functional assessment of cancer therapy-kidney symptom index-19. DRS=disease-related symptoms. DRS-P=disease-related symptoms-physical. DRS-E=disease-related symptoms-emotional. TSE-treatment side-effects. FWB=functional wellbeing. FACT-G=functional assessment of cancer therapy-general. PWB=physical wellbeing. EWB=emotional wellbeing. SWB=social and family wellbeing.

analysis, better PRO scores were consistently observed with nivolumab plus ipilimumab than with sunitinib in participants at intermediate or poor risk. Accordingly, analyses of changes from baseline over the first 103 weeks on study treatment showed improved HRQoL for nivolumab plus ipilimumab compared with sunitinib, with differences between treatment groups reaching significance at most timepoints for FKSI-19 and FACT-G total scores and most domain scores. Although a decrease in scores was observed at the first nivolumab plus ipilimumab induction visit, scores consistently recovered and thereafter improved and were maintained throughout the trial. This finding suggests that participants given nivolumab plus ipilimumab showed improved PROs even during the induction phase of combination treatment. In line with this observation, risk of deterioration in PRO scores was significantly decreased by nivolumab plus ipilimumab for the FKSI-19 and FACT-G total scores, and for the FKSI-19 disease-related symptoms, physical disease-related symptoms, treatment side-effects, and functional wellbeing domain scores, and the FACT-G physical wellbeing and functional wellbeing domain scores.

In all participants at intermediate or poor risk in the trial, we found that clinically meaningful improvements in FKSI-19 total and domain scores, FACT-G total score, and FACT-G physical wellbeing, emotional wellbeing, and functional wellbeing domain scores were prognostic of both progression-free survival and overall survival, although, notably, clinically important thresholds for change for the FKSI-19 instrument have not been fully validated in the literature for the total score and every domain score. We completed a post-hoc analysis of associations between baseline FKSI-19 total score, improvement in FKSI-19 total score, and overall survival. In this analysis, we pooled the nivolumab plus ipilimumab and sunitinib groups to determine if PROs are associated with clinical outcomes like survival, irrespective of the treatment received, and therefore, hold clinical relevance. We found that patients with high baseline HRQoL scores and HRQoL that improved from baseline had the longest overall survival. Similar results have been reported for other cancer types, including advanced lung,[25] cervical,[26] and ovarian cancers.[27] This observation suggests that baseline HRQoL and change in HRQoL, both improvement and worsening, has clinical prognostic value on survival outcomes. Furthermore, in an another post-hoc analysis of change in PRO scores from baseline for participants with high and low scores at baseline, similar numbers of participants had available PRO data later in the course of treatment in both these two groups. We therefore do not believe that results of the analysis for association of PROs with survival were subject to survival bias. Future research should examine how HRQoL information could inform treatment decision making in the advanced renal cell carcinoma treatment setting.

As has been previously reported,[11] in the treated population, treatment-related adverse events led to discontinuation of treatment for 118 (22%) of 547 participants in the nivolumab plus ipilimumab group and 63 (12%) of 535 in the sunitinib group. This difference in discontinuation due to treatment-related adverse events could be partly due to the different discontinuation criteria for both regimens and a better knowledge of how to manage sunitinib toxicities than those of nivolumab plus ipilimumab.[11] Dose reductions were not permitted in the nivolumab plus ipilimumab group; instead only dose delays were permitted. In the sunitinib group, patients were allowed to have up to two dose reductions before permanent discontinuation of treatment due to drug-related toxicity. In the combination group, as previously reported,[11] 319 (58%) of 547 patients had dose delays for nivolumab and 148 (27%) had dose delays for ipilimumab. In the sunitinib group, 315 (59%) of 535 patients had dose delays and 283 (53%) had dose reductions. Although between the nivolumab plus ipilimumab and sunitinib groups the overall proportions of participants who discontinued treatment (419 [77%] of 547 ipilimumab and nivolumab *vs* 438 [82%] of 535 sunitinib) and the overall treatment duration (median 7·9 months [95% CI 6·5–8·4] *vs* 7·8 months [6·4–8·5], respectively)[11] were similar, our results show that nivolumab plus ipilimumab improved PROs and HRQoL compared with sunitinib in CheckMate 214, supporting the positive overall benefit-risk profile for the combination treatment. We believe these results will be important in clinical treatment decisions for patients with advanced renal cell carcinoma.

A key strength of our analysis was the evaluation of the treatment effect on changes in PROs using prespecified thresholds to define deterioration in a randomised clinical trial with a large study sample. Additionally, we used both a generic instrument designed for the general population (EQ-5D-3L[28]) and cancer-specific instruments (FACT-G[13] and FKSI-19[17]) to measure HRQoL. Consistent significant improvements with nivolumab plus ipilimumab over sunitinib were observed for the FACT-G and FSKI-19 instruments and the EQ-5D-3L utility index, but not the EQ-5D-3L VAS. Given its generic nature, EQ-5D-3L might not detect small changes in health that are important for HRQoL studies in patients with cancer. Other studies support our findings. For example, in a phase 3, randomised, active-controlled trial of patients with advanced renal cell carcinoma by Castellano and collegues,[29] sunitinib led to significantly and clinically meaningful changes in FKSI-19 and FACT-G scores compared with interferon-γ, but the changes in EQ-5D-3L VAS did not reach significance. Similarly, in the phase 3, randomised, placebo-controlled VEG105192 trial in patients with advanced renal cell carcinoma,[23] EQ-5D-3L scores did not differ between patients given pazopanib and those given placebo when analysed by longitudinal MMRM analysis over 48 weeks. We believe that future trials in advanced renal cell carcinoma could preferentially use the FKSI-19 instrument, with the FACT-G and EQ-5D-3L also used depending on the interests of the investigating team.

Comparison of the baseline scores for the three PRO instruments in CheckMate 214 with baseline scores in other studies of first-line treatments for advanced renal cell carcinoma is difficult because few data are available for patients with intermediate or poor risk. The phase 2 CABOSUN trial[6] of cabozantinib versus sunitinib in patients at intermediate or poor risk did not collect PROs. Full PRO data from IMmotion151—a phase 3 trial of first-line atezolizumab plus bevacizumab in patients with renal cell carcinoma of any risk—have not yet been published, but improvement from baseline for FKSI-19 total scores with atezolizumab plus bevacizumab compared with sunitinib was reported at the 2018 American Society of Clinical Oncology Annual Meeting.[30,31]

Between-drug comparisons of HRQoL associated with targeted therapies in the first-line setting are scarce.[5] In the phase 3 COMPARZ trial of pazopanib versus sunitinib for first-line treatment of metastatic renal cell carcinoma, pazopanib showed similar efficacy but a more favourable safety profile and better HRQoL scores than sunitinib.[5] Further studies are necessary to better characterise differences in HRQoL with the growing treatment armamentarium available to patients with advanced renal cell carcinoma.

Our trial had some limitations that should be considered when interpreting the results. One such limitation is its open-label nature, which is considered a potential cause of bias. However, the sunitinib control group was an active therapy, which was considered the standard of care and had a more favourable administration route than nivolumab plus ipilimumab. And, despite the open-label design, we found that baseline and longitudinal changes in PROs in CheckMate 214 were prognostic of survival irrespective of treatment group, and consistent with previously PRO literature in other tumour types.[25–27] Therefore, although we cannot rule out the potential bias caused by the open-label design, we do not believe it meaningfully affected the PROs measured in this trial. Furthermore, although the different administration routes and schedules between the treatment groups could have influenced the results, we believe this effect would be mitigated by the timing of the PRO assessments relative to the dosing schedule. When PRO assessments were administered on day 1 of week 1 of the first two cycles, participants in the nivolumab plus ipilimumab group had been given their latest dose during the previous week, while those in the sunitinib group were being given sunitinib daily. When assessments were given on day 1 of week 5 of subsequent cycles, patients in the nivolumab plus ipilimumab group had been given their previous nivolumab dose 2 weeks before, and patients in the sunitinib group had been off treatment for a week before the assessment, so toxicities due to sunitinib would be less likely to affect PROs at these timepoints than when

the patient was receiving daily sunitinib. The different intervals between treatment groups from when the last dose was received to when PROs were assessed is also in line with the disparate half-lives of the different regimens.[8,9,32] Therefore, we believe that any potential bias introduced by the timing of the latest dose relative to when PROs were assessed was balanced between treatment groups. Several PRO measures were collected at different timepoints and multiplicity could be an issue; hence, we used repeated-measures analyses to adjust for this difference over time. Furthermore, in the association analyses, since a single hypothesis was not being tested, but rather several hypotheses about PRO measures explored individually, issues with multiplicity did not arise with regards to those individual hypotheses. However, we still proceeded with caution when drawing a single general conclusion about the association between PROs and the endpoints of interest when any of the multiple tests of the individual PRO measures are used as evidence of this general association between PRO and the endpoints.

Another limitation is the progressive decrease in the number of patients with PRO data throughout the trial. This decrease is because patients who discontinued treatment (mainly triggered by disease progression) had only two follow-up visits, after which all PRO assessments were discontinued except for EQ-5D-3L. The absence of PRO data after treatment discontinuation is a well-established drawback of clinical studies that incorporate PROs as secondary endpoints. Although collection of PRO data after treatment discontinuation could provide valuable insight into disease progression and HRQoL, subsequent treatments taken after discontinuation of study drug would make the interpretation of results challenging. The estimand for our change from baseline analysis is the so-called *de jure* estimand and corresponds to the while-on-treatment strategy. This approach means that we investigated the population-average effect, assuming all patients remained on the initial randomised treatment throughout the trial.[33] Our sensitivity analysis with PMM, which assumed that missing data were not missing at random, partly addressed the concern regarding the small number of patients still on treatment by week 103. By contrast, the time-to-deterioration analysis included all available assessments, including the two follow-up visits that took place after discontinuation of the study drug and therefore reflected the PRO scores beyond treatment discontinuation.

In conclusion, these PROs of HRQoL assessments from the primary analysis CheckMate 214 show that nivolumab plus ipilimumab combination treatment was associated with sustained improvement in all HRQoL measures compared with sunitinib monotherapy in previously untreated patients at intermediate or poor risk with advanced renal cell carcinoma. These results lend further support to the benefit-risk profile of nivolumab plus ipilimumab over sunitinib as a new standard of care for first-line, advanced renal cell carcinoma. Extended follow-up of CheckMate 214 is ongoing to further substantiate the long-term benefits of nivolumab plus ipilimuab in this setting.

Contributors

DC, JD, and RJM contributed to the conception and design of the study. VG, BE, HJH, SG, PN, M-OG, BIR, and RJM provided study materials or patients. CI and JP completed the statistical analyses. SM reviewed the clinical data. All authors contributed to the data analysis and interpretation, and drafting and revising of the manuscript.

Declaration of interests

DC reports grants and personal fees from Bristol-Myers Squibb during the conduct of the study and outside of the submitted work, and grants and personal fees from Pfizer, Novartis, Exelixis, Genentech Roche, Ipsen, Bayer, and AstraZeneca outside of the submitted work. VG reports honoraria from AstraZeneca, Bristol-Myers Squibb, Eisai, Ipsen, MSD, Merck Serono, Novartis, Pfizer, and Roche; has advisory roles with AstraZeneca, Bristol-Myers Squibb, Ipsen, MSD, Merck Serono, Novartis, Pfizer, and Roche; and research grants from AstraZeneca, Bristol-Myers Squibb, MSD, Novartis, and Pfizer. BE reports personal fees from Bristol-Myers Squibb during the conduct of the study, and personal fees from Bayer, Novartis, Pfizer, Exelixis, and Roche outside of the submitted work. HJH reports grants and personal fees from Bristol-Myers Squibb and Merck; personal fees from Pfizer and Exelixis; and grants and personal fees from Merck, during the conduct of the study. SG reports personal fees from AstraZeneca, Exelixis, Janssen, Genentech, and Sanofi Genzyme; grants and personal fees from Bayer, Bristol-Myers Squibb, Novartis, Corvus, and Pfizer; and grants from Acceleron, Merck, Agensys, and Eisai outside of the submitted work. PN reports personal fees from Bristol-Myers Squibb during the conduct of the study. M-OG reports grants and personal fees from Novartis and Bristol-Myers Squibb, and institutional support from Bristol-Myers Squibb for participation in the CheckMate 214 trial; and personal fees from Pfizer, Bayer HealthCare, Astellas, Intuitive Surgical, Sanofi Aventis, Hexal, Apogepha, Amgen, AstraZeneca, MSD, Janssen Cilag, ONO Pharmaceutical, and Ipsen Pharma outside of the submitted work. BIR reports grants and personal fees from Bristol-Myers Squibb outside of the submitted work. JD is an employee of Bristol-Myers Squibb. CI reports consulting fees from Bristol-Myers Squibb. JP reports consulting fees from Bristol-Myers Squibb. SM is an employee of Bristol-Myers Squibb. RJM reports grants from Bristol-Myers Squibb paid to his employer during the conduct of the study; grants and personal fees from Pfizer, Novartis, Eisai, Exelixis, Genentech Roche, Eisai, and Pfizer; and personal fees from Novartis outside of the submitted work.

Data sharing

Bristol-Myers Squibb's policy on data sharing can be found at https://www.bms.com/researchers-and-partners/independent-research/data-sharing-request-process.html. Deidentified and anonymised datasets of clinical trial information, including patient-level data, will be shared with external researchers for proposals that are complete, for which the scientific request is valid and the data are available, consistent with safeguarding patient privacy and informed consent. Upon execution of an agreement, the deidentified and anonymised data sets can be accessed via a secured portal that provides an environment for statistical programming with R as the programming language. The patient-reported outcomes protocol and statistical analysis plan will also be available. Data will be available for 2 years from the study completion or termination of the program (August, 2024).

Acknowledgments

We thank the patients, their families, and the investigators and participating study teams. Patients treated at Memorial Sloan Kettering Cancer Center were supported in part by Memorial Sloan Kettering Cancer Center Support Grant/Core Grant (P30 CA008748). Professional medical writing and editing assistance was provided by Nicolette Belletier, and Lawrence Hargett of PPSI (a PAREXEL company), and Montserrat Casamayor of IQVIA, funded by Bristol-Myers Squibb.

References

1 Motzer RJ, Escudier B, McDermott DF, et al. Nivolumab versus everolimus in advanced renal-cell carcinoma. *N Engl J Med* 2015; **373:** 1803–13.

2 Choueiri TK, Motzer RJ. Systemic therapy for metastatic renal-cell carcinoma. *N Engl J Med* 2017; **376:** 354–66.

3 Heng DY, Xie W, Regan MM, et al. External validation and comparison with other models of the International Metastatic Renal-Cell Carcinoma Database Consortium prognostic model: a population-based study. *Lancet Oncol* 2013; **14:** 141–48.

4 Heng DY, Xie W, Regan MM, et al. Prognostic factors for overall survival in patients with metastatic renal cell carcinoma treated with vascular endothelial growth factor-targeted agents: results from a large, multicenter study. *J Clin Oncol* 2009; **27:** 5794–99.

5 Motzer RJ, Hutson TE, Cella D, et al. Pazopanib versus sunitinib in metastatic renal-cell carcinoma. *N Engl J Med* 2013; **369:** 722–31.

6 Choueiri TK, Halabi S, Sanford BL, et al. Cabozantinib versus sunitinib as initial targeted therapy for patients with metastatic renal cell carcinoma of poor or intermediate risk: The Alliance A031203 CABOSUN trial. *J Clin Oncol* 2017; **35:** 591–97.

7 Cella D. Beyond traditional outcomes: improving quality of life in patients with renal cell carcinoma. *Oncologist* 2011; **16** (suppl 2): 23–31.

8 Opdivo (nivolumab) injection, for intravenous use. Princeton, NJ: Bristol-Myers Squibb Company, 2018. https://packageinserts.bms.com/pi/pi_opdivo.pdf (accessed Jan 7, 2019)

9 Yervoy (ipilimumab) injection, for intravenous use. Princeton, NJ; Bristol-Myers Squibb, 2018. https://packageinserts.bms.com/pi/pi_yervoy.pdf (accessed Jan 7, 2019).

10 Wolchok JD, Chiarion-Sileni V, Gonzalez R, et al. Overall survival with combined nivolumab and ipilimumab in advanced melanoma. *N Engl J Med* 2017; **377:** 1345–56.

11 Motzer RJ, Tannir NM, McDermott DF, et al. Nivolumab plus ipilimumab versus sunitinib in advanced renal-cell carcinoma. *N Engl J Med* 2018; **378:** 1277–90.

12 Rothrock NE, Jensen SE, Beaumont JL, et al. Development and initial validation of the NCCN/FACT symptom index for advanced kidney cancer. *Value Health* 2013; **16:** 789–96.

13 Cella DF, Tulsky DS, Gray G, et al. The Functional Assessment of Cancer Therapy scale: development and validation of the general measure. *J Clin Oncol* 1993; **11:** 570–79.

14 Euro-Qol Group. EuroQol—a new facility for the measurement of health-related quality of life. *Health Policy* 1990; **16:** 199–208.

15 Rao D, Butt Z, Rosenbloom S, et al. A Comparison of the Renal Cell Carcinoma-Symptom Index (RCC-SI) and the Functional Assessment of Cancer Therapy-Kidney Symptom Index (FKSI). *J Pain Symptom Manage* 2009; **38**: 291–98.

16 Dolan P. Modeling valuations for EuroQol health states. *Med Care* 1997; **35:** 1095–108.

17 Cella D, Yount S, Brucker PS, et al. Development and validation of a scale to measure disease-related symptoms of kidney cancer. *Value Health* 2007; **10:** 285–93.

18 Yost KJ, Eton DT. Combining distribution- and anchor-based approaches to determine minimally important differences: the FACIT experience. *Eval Health Prof* 2005; **28:** 172–91.

19 Pickard AS, Neary MP, Cella D. Estimation of minimally important differences in EQ-5D utility and VAS scores in cancer. *Health Qual Life Outcomes* 2007; **5:** 70.

20 O'Kelly M, Ratitch B. Clinical trials with missing data: a guide for practitioners. West Sussex: John Wiley & Sons, 2014.

21 Fairclough DL. Design and analysis of quality of life studies in clinical trials, 2nd edn. Boca Raton, FL: Chapman & Hall, CRC, 2010.

22 Motzer RJ, Hutson TE, Tomczak P, et al. Sunitinib versus interferon alfa in metastatic renal-cell carcinoma. *N Engl J Med* 2007; **356:** 115–24.

23 Sternberg CN, Davis ID, Mardiak J, et al. Pazopanib in locally advanced or metastatic renal cell carcinoma: results of a randomized phase III trial. *J Clin Oncol* 2010; **28:** 1061–68.

24 Cella D, Escudier B, Rini B, et al. Patient-reported outcomes for axitinib vs sorafenib in metastatic renal cell carcinoma: phase III (AXIS) trial. *Br J Cancer* 2013; **108:** 1571–78.

25 Eton DT, Fairclough DL, Cella D, et al. Early change in patient-reported health during lung cancer chemotherapy predicts clinical outcomes beyond those predicted by baseline report: results from Eastern Cooperative Oncology Group Study 5592. *J Clin Oncol* 2003; **21:** 1536–43.

26 Chase DM, Huang HQ, Wenzel L, et al. Quality of life and survival in advanced cervical cancer: a Gynecologic Oncology Group study. *Gynecol Oncol* 2012; **125:** 315–19.

27 Phippen NT, Secord AA, Wolf S, et al. Quality of life is significantly associated with survival in women with advanced epithelial ovarian cancer: an ancillary data analysis of the NRG Oncology/Gynecologic Oncology Group (GOG-0218) study. *Gynecol Oncol* 2017; **147:** 98–103.

28 Schrag A, Selai C, Jahanshahi M, Quinn NP. The EQ-5D—a generic quality of life measure—is a useful instrument to measure quality of life in patients with Parkinson's disease. *J Neurol Neurosurg Psychiatry* 2000; **69:** 67–73.

29 Castellano D, del Muro XG, Pérez-Gracia JL, et al. Patient-reported outcomes in a phase III, randomized study of sunitinib versus interferon-α as first-line systemic therapy for patients with metastatic renal cell carcinoma in a European population. *Ann Oncol* 2009; **20:** 1803–12.

30 Motzer RJ, Powles T, Atkins MB, et al. IMmotion151: a randomized phase III study of atezolizumab plus bevacizumab *vs* sunitinib in untreated metastatic renal cell carcinoma (mRCC). *Proc Am Soc Clin Oncol* 2018; **36** (suppl 6): 578 (abstr).

31 Escudier B, Motzer RJ, Rini B, et al. Patient-reported outcomes (PROs) in IMmotion151: atezolizumab (atezo) + bevacizumab (bev) vs sunitinib (sun) in treatment (tx) naive metastatic renal cell carcinoma (mRCC). *J Clin Oncol* 2018; **36** (suppl): 4511 (abstr).

32 Sutent (sunitinib malate) capsules, oral. New York, NY: Pfizer Labs, 2011. https://www.accessdata.fda.gov/drugsatfda_docs/label/2011/021938s13s17s18lbl.pdf (accessed Jan 8, 2019).

33 Phillips A, Abellan-Andres J, Soren A, et al. Estimands: discussion points from the PSI estimands and sensitivity expert group. *Pharm Stat* 2017; **16:** 6–11.