Article title:

Real-World Comparative Effectiveness and Safety of Immune Checkpoint Inhibitors in Advanced Renal Cell Carcinoma: A Systematic Review

Journal name:

Targeted Oncology

Author information:

Chin Hang Yiu,[1,2] Gwyneth Yuet Yin Leung,[1] Isabelle C. Lee,[1] Christine Y. Lu (corresponding author)[1,2,3]

Affiliations:

[1] The University of Sydney School of Pharmacy, Camperdown, New South Wales, Australia

[2] Kolling Institute, Faculty of Medicine and Health, The University of Sydney and the Northern Sydney Local Health District, Sydney, New South Wales, Australia

[3] Department of Pharmacy, Royal North Shore Hospital, St Leonards, New South Wales, Australia

Corresponding author:

Prof. Christine Lu

christine.lu@sydney.edu.au

Supplementary Table S1: Changes to PROSPERO protocol with reasons

Number	Protocol	Change(s)	Reason(s)
1	The search covered studies from database inception to September 16, 2025.	The search was updated to include studies from database inception to January 21, 2026.	To ensure that the search is up to date, no newly published articles are missing.
2	Search terms did not include individual ICI drug names.	Search terms were expanded to include all 15 USFDA-approved ICI agents, regardless of whether approval was specific to RCC.	To improve search sensitivity and ensure comprehensive identification of relevant real-world studies reporting ICI use in advanced RCC.

Supplementary Table S2: Full search strategy

	MEDLINE
1	Carcinoma, Renal Cell/
2	"renal cell carcinoma".mp.
3	1 or 2
4	Immune Checkpoint Inhibitors/
5	("immune checkpoint inhibitor*" or immunotherap* or "Atezolizumab" or "Durvalumab" or "Avelumab" or "Cosibelimab" or "Nivolumab" or "Pembrolizumab" or "Cemiplimab" or "Dostarlimab" or "Retifanlimab" or "Toripalimab" or "Tislelizumab" or "Penpulimab" or "Ipilimumab" or "Tremelimumab" or "Relatlimab").mp.
6	4 or 5
7	Administrative Claims, Healthcare/
8	Medical Records/
9	Electronic Health Records/
10	("claim*" or "administrative claim*" or "routine clinical" or "hospital" or "administrative data" or "claim* data*" or "prescription data*" or "health*data*" or "health system*" or "medical record*" or "medical chart*" or "electronic health record*" or "real-world").mp.
11	7 or 8 or 9 or 10
12	3 and 6 and 11
	Embase
1	renal cell carcinoma/
2	"renal cell carcinoma".mp.
3	1 or 2
4	immune checkpoint inhibitor/
5	("immune checkpoint inhibitor*" or immunotherap* or "Atezolizumab" or "Durvalumab" or "Avelumab" or "Cosibelimab" or "Nivolumab" or "Pembrolizumab" or "Cemiplimab" or "Dostarlimab" or "Retifanlimab" or "Toripalimab" or "Tislelizumab" or "Penpulimab" or "Ipilimumab" or "Tremelimumab" or "Relatlimab").mp.
6	4 or 5
7	"administrative claims (health care)"/
8	medical record/
9	electronic health record/
10	("claim*" or "administrative claim*" or "routine clinical" or "hospital" or "administrative data" or "claim* data*" or "prescription data*" or "health*data*" or "health system*" or "medical record*" or "medical chart*" or "electronic health record*" or "real-world").mp.
11	7 or 8 or 9 or 10
12	3 and 6 and 11
	Scopus
(TITLE-ABS-KEY ("renal cell carcinoma") AND TITLE-ABS-KEY (("immune checkpoint inhibitor*" OR immunotherap* OR "Atezolizumab" or "Durvalumab" OR "Avelumab" or "Cosibelimab" OR "Nivolumab" OR "Pembrolizumab" OR "Cemiplimab" OR "Dostarlimab" OR "Retifanlimab" OR "Toripalimab" OR "Tislelizumab" OR "Penpulimab" OR "Ipilimumab" OR "Tremelimumab" OR "Relatlimab")) AND TITLE-ABS-KEY (("claim*" OR "electronic health record*" OR "medical record*" OR "medical chart*" OR "real-world")))	

Supplementary Table S3: Quality assessment for studies reporting efficacy outcomes (n = 57 studies)

	Selection domain				Comparability domain	Outcome domain			Overall Quality
Study	**Representativeness of exposed cohort (Max: 1★)**	**Selection of the non-exposed cohort (Max: 1★)**	**Ascertainment of exposure (Max: 1★)**	**Demonstration that outcome of interest was not present at start of study (Max: 1★)**	**Comparability of cohorts on the basis of design and analysis (Max: 2★)**	**Assessment of outcome (Max: 1★)**	**Was follow-up long enough for outcomes to occur (Max: 1★)**	**Adequacy of follow-up cohorts (Max: 1★)**	
De Vries-Brilland et al. (2025) [29]	★	★	★	–	★ Cox models adjusted only for IMDC risk. No adjustment for other variables such as age and sex.	★	★ Median follow-up 21.6 months	– Proportion lost to follow-up not clearly quantified	Good
Massari et al. (2025) [30]	★	★	★	–	– Strong baseline imbalances and no adjustment for treatment differences. While multivariable Cox regression was used to identify prognostic factors, treatment regimens were not included.	★	★ Median follow-up 21.2 months	– Proportion lost to follow-up not clearly quantified	Poor
Hilser et al. (2025) [31]	★	★	★	–	– Unadjusted Kaplan-Meier analyses for treatment comparisons.	★	★ Median follow-up 33.3 months	– Proportion lost to follow-up not clearly quantified	Poor
Ostrowski et al. (2025) [32]	★	★	★	–	★★ Propensity score matching–weighted analysis. The model includes: age, race/ethnicity, smoking, insurance, practice type, prior nephrectomy, IMDC risk, year of treatment initiation.	★	★ Median follow-up 21.16 months	– Proportion lost to follow-up not clearly quantified	Good

Haack et al. (2025) [33]	★	★	★	–	★ Multivariable Cox regression includes treatment group (IO-TKI vs IO-IO) in the PFS analysis. Adjustment includes limited covariates (risk factors such as synchronous metastases, IMDC, selected clinical features)	★	★ Median follow-up 21 months	– Proportion lost to follow-up not clearly quantified	Good
Ishihara et al. (2025) [34]	★	★	★	–	★★ Inverse Probability of Treatment Weighting (IPTW) analysis to balance the characteristics of the cohorts. It achieved negligible imbalances (Standardized Mean Difference < 10%) across critical factors: age, sex, previous nephrectomy, histopathology, metastatic status, IMDC risk, and site-specific metastases.	★	★ Median follow-up 17.4 months	– Proportion lost to follow-up not clearly quantified	Good
Santoni et al. (2025) [35]	★	★	★	–	★ Multivariable Cox regression includes treatment groups for OS and PFS analyses. Adjustment includes several important prognostic factors such as age, IMDC, nephrectomy, sarcomatoid features.	★	★ Median follow-up 17 months	– Proportion lost to follow-up not clearly quantified	Good
Gupta et al. (2025) [36]	★	★	★	–	– No formal statistical testing or matching/weighting was undertaken to compare survival periods between treatment subgroups. While stratified by IMDC risk, significant baseline imbalances remained in age, sites of metastasis, and prior nephrectomy.	★	★ Median follow-up 33.2 months	– Proportion lost to follow-up not clearly quantified	Poor

Kikuta et al. (2025) [37]	★	★	★	–	★★ Inverse Probability of Treatment Weighting (IPTW) analysis to balance the important characteristics of the cohorts such as age, sex, sacromatoid features, IMDC, prior nephrectomy, site-specific metastases.	★	★ Median follow-up 15.8 months	– Proportion lost to follow-up not clearly quantified	Good
Hara et al. (2025) [38]	★	★	★	–	★★ Propensity score–based IPTW performed across critical factors such as age, BMI, tumor size, nephrectomy status, histopathology, and IMDC risk. Excellent balance achieved (all SMDs < 0.10 after weighting).	★	★ Median follow-up 26.5 months (ICI-ICI) and 12.2 (ICI -TKI) months.	– Proportion lost to follow-up not clearly quantified	Good
Yanagisawa et al. (2024) [39]	★	★	★	–	★★ Propensity score–matched analysis performed based on IMDC risk classification, with excellent post-matching balance achieved (all standardized mean differences < 0.10). Treatment groups were well balanced in IMDC risk distribution and major baseline clinical characteristics (e.g. age, sex, metastatic sites) after matching. (Note: matching was limited to IMDC risk only; no additional covariates were included in the PS model.)	★	★ Median follow-up 17 months	– Proportion lost to follow-up not clearly quantified	Good

Shah et al. (2024) [40]	★	★	★	–	★ Multivariable regression adjustment performed across key baseline factors including age, sex, insurance type, geographic region, comorbidity burden (Charlson index), number of metastatic sites, metastatic site location, and index year. No propensity score–based matching or weighting and no balance diagnostics reported.	★	★ Median follow-up 8.6 months	– Proportion lost to follow-up not clearly quantified	Good
Roviello et al. (2025) [41]	★	★	★	–	– Unadjusted Kaplan-Meier analyses for treatment comparisons.	★	★ Median follow-up 37.2 months	– Proportion lost to follow-up not clearly quantified	Poor
Ishikawa et al. (2025) [42]	★	★	★	–	– While baseline characteristics were balanced, the study used unadjusted Kaplan-Meier estimates for primary survival comparisons. Multivariable Cox regression was used to identify independent prognostic factors, but it did not provide a balanced treatment-effect adjustment	★	★ Median follow-up 66 months (TKI) and 23 (ICI-TKI) months.	– Proportion lost to follow-up not clearly quantified	Poor
Ciccarese et al. (2024) [43]	★	★	★	–	– Only univariate comparisons were conducted different treatment groups.	★	★ Median follow-up 15.6 months	– Proportion lost to follow-up not clearly quantified	Poor
Yoshimura et al. (2024) [44]	★	★	★	–	– Univariate Kaplan–Meier analyses were used to compare treatment groups, while multivariable Cox regression was conducted only for prognostic factor assessment (e.g., sex and liver metastasis) and did not include treatment group as an adjusted covariate.	★	★ Median follow-up 15.2 months	– Proportion lost to follow-up not clearly quantified	Poor

Bickley et al. (2024) [45]	★	★	★	–	★ Multivariable Cox proportional hazards models included treatment group and adjusted for key prognostic factors such as IMDC, metastatic sites.	★	★ Although median follow-up was not reported, the study period (2015–2022) and data curation likely provided enough longitudinal visibility.	– Proportion lost to follow-up not clearly quantified	Good
Liu et al. (2024) [46]	★	★	★	–	– No formal treatment-effect adjustment was performed.	★	★ Median follow-up two years	– Proportion lost to follow-up not clearly quantified	Poor
Massari et al. (2024) [47]	★	★	★	–	★ Multivariable Cox proportional hazards models included treatment group and adjusted for key prognostic factors such as bone metastases, liver metastases, IMDC risk group, nephrectomy status, age, and sex.	★	★ Median follow-up 22.9 months	– Proportion lost to follow-up not clearly quantified	Good
Gebrael et al. (2024) [48]	★	★	★	–	★★ Propensity score–based matching weighting performed using key baseline covariates including second-line start year, age, race/ethnicity, practice type, insurance status, IMDC risk group, and prior nephrectomy. Excellent covariate balance achieved after weighting (all standardized mean differences < 0.10).	★	★ Although median follow-up was not reported, the study period (2017–2022) and data curation likely provided enough longitudinal visibility.	– Proportion lost to follow-up not clearly quantified	Good

Ishihara et al. (2024) [49]	★	★	★	–	★★ Inverse Probability of Treatment Weighting (IPTW) analysis to balance baseline characteristics, including age, sex, prior nephrectomy, histology, and IMDC risk factors. All covariates achieved excellent balance with a Standardized Mean Difference (SMD) < 0.1.	★	★ Median follow-up 15.0 months	– Proportion lost to follow-up not clearly quantified	Good
Albigès et al. (2023) [50]	★	★	★	–	– No formal treatment-effect adjustment was performed. Survival analyses were descriptive and stratified by treatment class only, using Kaplan–Meier methods.	★	★ Median follow-up 41.4 months (maximum 48 months follow-up)	– Proportion lost to follow-up not clearly quantified	Poor
Sawaya et al. (2024) [51]	★	★	★	–	★★ Inverse Probability of Treatment Weighting (IPTW) using propensity scores to balance cohorts. It achieved excellent balance (ASMD < 10%) across all baseline characteristics, including IMDC risk, histology, and sites of metastasis.	★	★ Median follow-up 21 months	– Proportion lost to follow-up not clearly quantified	Good
Huang et al. 2024 [52]	★	★	★	–	★★ Inverse Probability of Treatment Weighting (IPTW) method to balance patient baseline characteristics.	★	★ Median follow-up 27.3 months	– Proportion lost to follow-up not clearly quantified	Good
Esterberg et al. (2024) [53]	★	★	★	–	★ Multivariable Cox regression model to adjust for several key prognostic factors, including IMDC risk category, histology, gender, and sites of metastasis.	★	★ Median follow-up 18.7 months	– Proportion lost to follow-up not clearly quantified	Good

Santoni et al. (2024) [54]	★	★	★	–	– Primary survival comparisons between treatment groups were conducted using unadjusted Kaplan-Meier curves and log-rank tests. The multivariable Cox regression was used to identify independent prognostic factors (e.g., BMI, nephrectomy, bone/liver metastases).	★	★ Median follow-up 18.7 months	– Proportion lost to follow-up not clearly quantified	Poor
Lai et al. (2023) [55]	★	★	★	–	– Unadjusted Kaplan-Meier survival analyses and raw group comparisons. While baseline imbalances were noted (e.g., Ipi+Nivo cohort had more metastases and comorbidities), no formal multivariable adjustment or propensity-based balancing was performed to account for these differences.	★	★ Mean follow-up 35.3 months (TKI), 19.7 months (ICI-ICI), 16.7 months (ICI-TKI)	– Proportion lost to follow-up not clearly quantified	Poor
Poprach et al. (2023) [57]	★	★	★	–	★ A multivariable Cox proportional-hazards model to adjust for several key prognostic factors, including MSKCC risk score, tumor grade, bone/liver/lung/lymph node metastases, and prior nephrectomy.	★	★ Median follow-up 18.5 months (TT) and 29.3 months (IO) months	– Proportion lost to follow-up not clearly quantified	Good
Shah et al. (2023) [58]	★	★	★	–	★ A multivariable Cox proportional-hazards model to adjust for baseline clinical and demographic variables (age, gender, race, BMI, ECOG PS, IMDC risk).	★	★ Median follow-up 10.1 months (pembrolizumab-axitinib) and 10.7 months (ipilimumab-nivolumab) months	– Proportion lost to follow-up not clearly quantified	Good

Al-Mansour et al. (2024) [59]	★	★	★	–	– Unadjusted Kaplan-Meier analyses and Cox regression for treatment comparisons.	★	★ Median follow-up 16.5 months	– Proportion lost to follow-up not clearly quantified	Poor
Teishima et al. (2024) [60]	★	★	★	–	– No formal treatment-effect adjustment was performed. Comparisons between groups relied on unadjusted Kaplan–Meier analyses and log-rank tests.	★	★ Median follow-up 12.6 months	– Proportion lost to follow-up not clearly quantified	Poor
Shah et al. (2023) [61]	★	★	★	–	★ Multivariable Cox proportional hazards models included treatment group and adjusted for key prognostic factors such as age, sex, race, smoking status, BMI, ECOG performance status, histology, and IMDC risk group.	★	★ Median follow-up ~8 months	– Proportion lost to follow-up not clearly quantified	Good
Iinuma et al. (2023) [62]	★	★	★	–	– Unadjusted Kaplan-Meier methods and log-rank tests.While multivariable Cox regression was used to identify independent prognostic factors (e.g., age, poor IMDC risk), it did not provide a balanced treatment-effect adjustment.	★	★ Median follow-up 7 months	– Proportion lost to follow-up not clearly quantified	Poor
van Laar et al. (2022) [63]	★	★	★	–	– Unadjusted Kaplan-Meier survival analyses for the primary outcomes. No formal multivariable adjustment or propensity-based balancing was performed to account for the noted significant baseline imbalances in age, nephrectomy status, and risk categories between the treatment groups.	★	★ Although median follow-up was not reported, the study period (2015–2020) and data curation likely provided enough longitudinal visibility.	– Proportion lost to follow-up not clearly quantified	Poor

Kimura et al. (2020) [64]	★	★	★	–	– Unadjusted Kaplan-Meier survival curves and log-rank tests for OS. While baseline characteristics were compared and no significant differences were found, the study did not perform multivariable adjustment or propensity-based balancing to account for selection bias.	★	– Follow-up unclear	– Proportion lost to follow-up not clearly quantified	Poor
Ishihara et al. (2023) [65]	★	★	★	–	★ Multivariable Cox proportional-hazard regression to adjust for several key prognostic factors (age, histology, IMDC risk, metastatic sites).	★	★ Median follow-up 18.0 months	– Proportion lost to follow-up not clearly quantified	Good
Carril-Ajuria et al. (2023) [66]	★	★	★	–	– No formal treatment-effect adjustment was performed. Comparisons relied on unadjusted Kaplan–Meier analyses.	★	★ Median follow-up 11.1 months	– Proportion lost to follow-up not clearly quantified	Poor
Zarrabi et al. (2023) [67]	★	★	★	–	★★ Inverse Probability of Treatment Weighting (IPTW) analysis to balance groups across all pre-treatment covariates (age, gender, insurance, race, IMDC, and nephrectomy). Balance was achieved with Standardized Mean Differences (SMD) < 0.1.	★	★ Median follow-up 20.0 months	– Proportion lost to follow-up not clearly quantified	Good
Stühler et al. (2023) [68]	★	★	★	–	– No formal treatment-effect adjustment was performed. Comparisons relied on unadjusted Kaplan–Meier analyses. Only exploratory multivariate analysis was performed.	★	★ Median follow-up 29.9 months (TKI) and 24.1 months (ICI-based)	– Proportion lost to follow-up not clearly quantified	Poor

Santoni et al. (2022) [69]	★	★	★	–	– No formal treatment-effect adjustment was performed. Comparisons relied on unadjusted Kaplan–Meier analyses.	★	★ Median follow-up 57.4 months	– Proportion lost to follow-up not clearly quantified	Poor
Hoeh et al. (2022) [70]	★	★	★	–	– Unadjusted Kaplan-Meier analyses and univariate Cox regression for treatment comparisons. Treatment group was not included in a multivariable model	★	★ Median follow-up 9 months	– Proportion lost to follow-up not clearly quantified	Poor
Bando et al. (2022) [71]	★	★	★	–	– Unadjusted Kaplan-Meier analyses for treatment comparisons.	★	★ Mean follow-up 11.7 months	– Proportion lost to follow-up not clearly quantified	Poor
Bhanegaonkar et al. (2021) [72]	★	★	★	–	– Kaplan–Meier curves were presented within cohorts only. No statistical comparisons between treatment cohorts were performed	★	– Follow-up unclear	– Proportion lost to follow-up not clearly quantified	Poor
Kido et al. (2021) [73]	★	★	★	–	★★ Inverse Probability of Treatment Weighting (IPTW) to balance baseline covariates: age, sex, performance status, number of IMDC risk factors, history of nephrectomy, and number of metastatic organs.	★	★ Mean follow-up 12 (ICI-ICI) and 16 (TKI) months	– Proportion lost to follow-up not clearly quantified	Good
Stukalin et al. (2019) [74]	★	★	★	–	★ Multivariable Cox proportional hazards regression to adjust for several key prognostic factors, including the six criteria of the IMDC model and age.	★	– Follow-up unclear	– Proportion lost to follow-up not clearly quantified	Poor

McGrane et al. (2024) [75]	★	★	★	–	– Primary survival comparisons between treatment groups were conducted using unadjusted Kaplan-Meier curves and log-rank tests.	★	★ Median follow-up 16 months	– Proportion lost to follow-up not clearly quantified	Poor
Santoni et al. (2023) [76]	★	★	★	–	– Primary survival comparisons between treatment groups were conducted using unadjusted Kaplan-Meier curves and log-rank tests.	★	★ Median follow-up 18.1 months	– Proportion lost to follow-up not clearly quantified	Poor
Kato et al. (2022) [77]	★	★	★	–	★★ Inverse Probability of Treatment Weighting (IPTW) to balance cohorts across multiple clinical variables, including age, sex, IMDC risk, TNM stage, and site of metastases. Balance was achieved for most factors (standardized difference < 0.25).	★	★ Mean follow-up 23.6 (ICI-ICI) and 28.8 (TKI) months	– Proportion lost to follow-up not clearly quantified	Good
Ueda et al. (2022) [78]	★	★	★	–	★ Multivariable Cox proportional hazards modelling to adjust for several key prognostic factors (prior nephrectomy and performance status) in PFS outcome.	★	★ Median follow-up 21.4 months	– Proportion lost to follow-up not clearly quantified	Good
Bai et al. (2024) [79]	★	★	★	–	– Primary survival comparisons between treatment groups were conducted using unadjusted Kaplan-Meier curves and log-rank tests.	★	★ Median follow-up 34.5 months	– Proportion lost to follow-up not clearly quantified	Poor
Ueda et al. (2026) [80]	★	★	★	–	★ Multivariable Cox proportional hazards modelling to adjusting for ECOG performance status and hemodialysis duration.	★	★ Median follow-up 13.8 months	– Proportion lost to follow-up not clearly quantified	Good

Mukae et al. (2026) [81]	★	★	★	–	– Primary survival comparisons between treatment groups were conducted using unadjusted Kaplan-Meier curves and log-rank tests.	★	★ Median follow-up 17.5 months	– Proportion lost to follow-up not clearly quantified	Poor
Massari et al. (2025) [82]	★	★	★	–	– Primary survival comparisons between treatment groups were conducted using unadjusted Kaplan-Meier curves and log-rank tests.	★	★ Median follow-up 19.1 months	– Proportion lost to follow-up not clearly quantified	Poor
Geynisman et al. (2025) [83]	★	★	★	–	★★ Stabilized Inverse Probability of Treatment Weighting (IPTW) to balance cohorts across a wide range of baseline characteristics, including age, insurance, tumor grade, IMDC risk, and ECOG status. Balance was successfully achieved with Standardized Mean Differences (SMD) < 0.1 for all variables.	★	★ Median follow-up 22.4 (IO-IO) and 19.0 (IO-TKI) months	– Proportion lost to follow-up not clearly quantified	Good
Ikoma et al. (2025) [84]	★	★	★	–	– Primary survival comparisons between treatment groups were conducted using unadjusted Kaplan-Meier curves and log-rank tests.	★	★ Median follow-up 13.9 (older adults) and 25.9 (non-older adults) months	– Proportion lost to follow-up not clearly quantified	Poor
Meng et al. (2025) [85]	★	★	★	–	– Univariate Cox regression analysis to identify factors associated with overall survival.	★	★ Median follow-up 25.3 (TKI) and 17.3 (ICI) months	– Proportion lost to follow-up not clearly quantified	Poor

Goebell et al. (2025) [86]	★	★	★	–	★★ Stabilized Inverse Probability of Treatment Weighting (IPTW) to balance cohorts across a multitude of baseline variables (age, sex, IMDC, histology, comorbidities, metastatic sites). Balance was successfully achieved with Standardized Mean Differences (SMD) < 0.1 for all variables.	★	★ Patients were followed for up to 3 years	★ The study reports that only 5% of patients were lost to follow-up	Good

Good quality: 3 or 4 stars in selection domain AND 1 or 2 stars in comparability domain AND 2 or 3 stars in outcome domain

Fair quality: 2 stars in selection domain AND 1 or 2 stars in comparability domain AND 2 or 3 stars in outcome domain

Poor quality: 0 or 1 star in selection domain OR 0 stars in comparability domain OR 0 or 1 stars in outcome domain

Supplementary Table S4: Quality assessment for studies reporting safety outcomes (n = 24)

	Selection domain				Comparability domain	Outcome domain			Overall Quality
Study	**Representativeness of exposed cohort (Max: 1★)**	**Selection of the non-exposed cohort (Max: 1★)**	**Ascertainment of exposure (Max: 1★)**	**Demonstration that outcome of interest was not present at start of study (Max: 1★)**	**Comparability of cohorts on the basis of design and analysis (Max: 2★)**	**Assessment of outcome (Max: 1★)**	**Was follow-up long enough for outcomes to occur (Max: 1★)**	**Adequacy of follow-up cohorts (Max: 1★)**	
De Vries-Brilland et al. (2025) [29]	★	★	★	–	– No adjustment for treatment differences. Safety outcomes were reported descriptively and were not adjusted.	★	★ Median follow-up 21.6 months	★	Poor
Hilser et al. (2025) [31]	★	★	★	–	– No adjustment for treatment differences. Safety outcomes were reported descriptively and were not adjusted.	★	★ Median follow-up 21.6 months	★	Poor
Haack et al. (2025) [33]	★	★	★	–	– No adjustment for treatment differences. Safety outcomes were reported descriptively and were not adjusted.	★	★ Median follow-up 21 months	★	Poor
Ishihara et al. (2025) [34]	★	★	★	–	★★ Inverse Probability of Treatment Weighting (IPTW) analysis to balance the characteristics of the cohorts. It achieved negligible imbalances (Standardized Mean Difference < 10%) across critical factors: age, sex, previous nephrectomy, histopathology, metastatic status, IMDC risk, and site-specific metastases.	★	★ Median follow-up 17.4 months	★	Good
Yanagisawa et al. (2024) [39]	★	★	★	–	★★ Propensity score–matched analysis performed based on IMDC risk classification, with excellent post-matching balance achieved (all	★	★ Median follow-up 17 months	★	Good

					standardized mean differences < 0.10). Treatment groups were well balanced in IMDC risk distribution and major baseline clinical characteristics (e.g. age, sex, metastatic sites) after matching. (Note: matching was limited to IMDC risk only; no additional covariates were included in the PS model.)				
Shah et al. (2024) [40]	★	★	★	–	★ Multivariable regression adjustment performed across key baseline factors including age, sex, insurance type, geographic region, comorbidity burden (Charlson index), number of metastatic sites, metastatic site location, and index year. No propensity score–based matching or weighting and no balance diagnostics reported	– Health care resource use (HCRU) reflects all-cause utilisation, not treatment-attributable toxicity	★ Follow-up duration of median ~8.6 months is sufficient for capturing early HCRU events	★	Good
Ishikawa et al. (2025) [42]	★	★	★	–	– No adjustment for treatment differences. Safety outcomes were reported descriptively and were not adjusted.	★	★ Median follow-up 66 months (TKI) and 23 (ICI -TKI) months.	★	Poor
Yoshimura et al. (2024) [44]	★	★	★	–	– No adjustment for treatment differences. Safety outcomes were reported descriptively and were not adjusted.	★	★ Median follow-up 15.2 months	★	Poor
Liu et al. (2024) [46]	★	★	★	–	– No adjustment for treatment differences. Safety outcomes were reported descriptively and were not adjusted.	★	★ Median follow-up two years	★	Poor
Ishihara et al. (2024) [49]	★	★	★	–	★★ Inverse Probability of Treatment Weighting (IPTW) analysis to balance baseline characteristics,	★	★ Median follow-up 15.0 months	★	Good

					including age, sex, prior nephrectomy, histology, and IMDC risk factors. All covariates achieved excellent balance with a Standardized Mean Difference (SMD) < 0.1.				
Huang et al. 2024 [52]	★	★	★	–	– No adjustment for treatment differences. Safety outcomes were reported descriptively and were not adjusted (while IPTW was used for survival outcomes)	★	★ Median follow-up 27.3 months	★	Poor
Blas et al. (2024) [54]	★	★	★	–	– While logistic regression was used to calculate reporting odds ratios (RORs) for serious AEs, the analysis was not adjusted for key clinical confounders such as IMDC risk, age, or comorbidities.	– Assessment relied on voluntary and mandatory reporting to the FAERS database, which is prone to significant underreporting and variability in data completeness.	★ The analysis period (2013–2022) was sufficient to capture both early and late-onset toxicities associated with these regimens.	★	Poor
Teishima et al. (2024) [59]	★	★	★	–	– No adjustment for treatment differences. Safety outcomes were reported descriptively and were not adjusted.	★	★ Median follow-up 12.6 months	★	Poor
Kimura et al. (2020) [63]	★	★	★	–	– No adjustment for treatment differences. Safety outcomes were reported descriptively and were not adjusted.	★	– Follow-up unclear	★	Poor
Carril-Ajuria et al. (2023) [65]	★	★	★	–	– No adjustment for treatment differences. Safety outcomes were	★	★ Median follow-up 11.1 months	★	Poor

					reported descriptively and were not adjusted.				
Hoeh et al. (2022) [69]	★	★	★	–	– No adjustment for treatment differences. Safety outcomes were reported descriptively and were not adjusted.	★	– Median follow-up was relatively short.	★	Poor
Bando et al. (2022) [70]	★	★	★	–	– No adjustment for treatment differences. Safety outcomes were reported descriptively and were not adjusted.	★	★ Mean follow-up 11.7 months	★	Poor
Bhanegaonkar et al. (2021) [71]	★	★	★	–	– No adjustment for treatment differences. Safety outcomes were reported descriptively and were not adjusted.	★	– Follow-up unclear	★	Poor
Kido et al. (2021) [72]	★	★	★	–	– No adjustment for treatment differences. Safety outcomes were reported descriptively and were not adjusted (while IPTW was used for survival outcomes)	★	★ Mean follow-up 12 (ICI - ICI) and 16 (TKI) months	★	Poor
Kato et al. (2022) [76]	★	★	★	–	– No adjustment for treatment differences. Safety outcomes were reported descriptively and were not adjusted (while IPTW was used for survival outcomes)	★	★ Mean follow-up 23.6 (ICI - ICI) and 28.8 (TKI) months	★	Poor
Ueda et al. (2026) [79]	★	★	★	–	– No adjustment for treatment differences. Safety outcomes were reported descriptively and were not adjusted.	★	★ Median follow-up 13.8 months	★	Poor
Mukae et al. (2026) [80]	★	★	★	–	– No adjustment for treatment differences. Safety outcomes were	★	★ Median follow-up 17.5 months	★	Poor

					reported descriptively and were not adjusted.				
Geynisman et al. (2025) [83]	★	★	★	–	★★ Stabilized Inverse Probability of Treatment Weighting (IPTW) to balance cohorts across a wide range of baseline characteristics, including age, insurance, tumor grade, IMDC risk, and ECOG status. Balance was successfully achieved with Standardized Mean Differences (SMD) < 0.1 for all variables.	– Assessment relied on retrospective chart review where physicians could select only up to the 5 most "clinically significant" irAEs, potentially introducing reporting bias.	★ Median follow-up 22.4 (ICI-ICI) and 19.0 (ICI - TKI) months	★	Good
Ikoma et al. (2025) [84]	★	★	★	–	– No adjustment for treatment differences. Safety outcomes were reported descriptively and were not adjusted.	★	★ Median follow-up 13.9 (older adults) and 25.9 (non-older adults) months	★	Poor

Good quality: 3 or 4 stars in selection domain AND 1 or 2 stars in comparability domain AND 2 or 3 stars in outcome/exposure domain

Fair quality: 2 stars in selection domain AND 1 or 2 stars in comparability domain AND 2 or 3 stars in outcome/exposure domain

Poor quality: 0 or 1 star in selection domain OR 0 stars in comparability domain OR 0 or 1 stars in outcome/exposure domain