
Article:
Cooper, K. orcid.org/0000-0002-7702-8103, Tappenden, P., Cantrell, A. et al. (1 more author) (2020) A systematic review of meta-analyses assessing the validity of tumour response endpoints as surrogates for progression-free or overall survival in cancer. British Journal of Cancer. ISSN 0007-0920

https://doi.org/10.1038/s41416-020-01050-w

A systematic review of meta-analyses assessing the validity of tumour response endpoints as surrogates for progression-free or overall survival in cancer

Running title: Systematic review of response as surrogate for survival

List of authors:

*Katy Cooper, Senior Research Fellow, ScHARR, University of Sheffield, k.l.cooper@sheffield.ac.uk, ORCiD 0000-0002-7702-8103

Paul Tappenden, Professor of Health Economic Modelling, ScHARR, University of Sheffield

Anna Cantrell**,** Information Specialist, ScHARR, University of Sheffield

Kate Ennis, Research Associate, ScHARR, University of Sheffield

*corresponding author

25 February 2020

Competing interests

The authors have no competing interests.

Abstract

Background: Tumour response endpoints, such as overall response rate (ORR) and complete response (CR), are increasingly used in cancer trials. However, the validity of response-based surrogates is unclear. This systematic review summarises meta-analyses assessing the association between response-based outcomes and overall survival (OS), progression-free survival (PFS) or time-to-progression (TTP).

Methods: Five databases were searched to March 2019. Meta-analyses reporting correlation or regression between response-based outcomes and OS, PFS or TTP were summarised.

Results: The systematic review included 63 studies across 20 cancer types, most commonly non-small cell lung cancer (NSCLC), colorectal cancer (CRC) and breast cancer. The strength of association between ORR or CR and either PFS or OS varied widely between and within studies, with no clear pattern by cancer type. The association between ORR and OS appeared weaker and more variable than that between ORR and PFS, both for associations between absolute endpoints and associations between treatment effects.

Conclusions: This systematic review suggests that response-based endpoints such as ORR and CR may not be reliable surrogates for PFS or OS. Where it is necessary to use tumour response to predict treatment effects on survival outcomes, it is important to fully reflect all statistical uncertainty in the surrogate relationship.

Introduction

Decisions about the use of new and existing health technologies should ideally be informed by estimates of treatment effects derived from high quality randomised controlled trials (RCTs) which measure patient-relevant endpoints over a clinically appropriate timeframe. Such "final" endpoints typically involve the measurement of health benefits which reflect aspects of the disease and its treatment which are important to patients (and potentially also their carers) and which relate to "how the patient feels, functions or survives."[1] In the context of advanced/metastatic cancer, the key matter of concern is often whether the use of a given heath technology leads to improvements in overall survival (OS; a final endpoint) compared to existing standard treatments. However, the estimation of treatment effects on OS may be subject to numerous problems, including: potential confounding resulting from the use of post-progression treatments, insufficient study follow-up resulting in data immaturity, or simply that data on OS have not been collected. In such instances, determining the impact of health technologies becomes more challenging and may rely on the use of surrogate endpoints to substitute for, and predict, a final patient-relevant clinical outcome.[2] Potentially relevant surrogate endpoints vary according to tumour type and site, but commonly include progression-free survival (PFS), time to progression (TTP), and response-based outcomes, which may include overall response rate (ORR), different levels of response (e.g. complete response [CR], partial response [PR] or very good partial response [VGPR]) and duration of response (DoR). These surrogate endpoints are often considered attractive as they typically require smaller sample sizes, occur faster and are less expensive to collect in clinical trials compared with final outcomes, thereby reducing costs associated with data collection and expediting the time required for bringing new technologies to market.

It has been recognised in the literature that the reliance on surrogates may lead to invalid conclusions regarding the net health effects of technologies, which in turn, have the potential to lead to patient harm.[3] Much of the published literature around the use of surrogate endpoints has focussed on the development and application of frameworks for their validation.[4,5] In his seminal paper, Prentice[4] put forward stringent criteria for the validation of surrogate endpoints in Phase III trials. In general terms, these criteria require that the surrogate endpoint must be a correlate of the net effect of treatment on the

final clinical outcome – in other words, there must be a single pathway from the treatment to the true endpoint which is mediated exclusively by the surrogate endpoint.[6] Applied surrogate validation studies commonly adopt a meta-analytic (meta-regression) approach based on multiple studies in order to assess whether the apparent relationship between the surrogate and the final endpoint remains constant in the presence of various sources of heterogeneity, such as differences in patient population, study design and treatments received.[5]

Based on the NIH Biomarkers Definition Working Group's preferred terms and definitions[7] and the 2001 Journal of the American Medical Association (JAMA) User's Guide,[8] Taylor and Elston[9] proposed a hierarchy of levels of surrogate validation. Level 3 of the hierarchy relates to biological plausibility – this is the weakest form of validation and is typically based on pathophysiological studies and/or an understanding of the disease process. Level 2 requires the presence of a consistent association between the surrogate outcome and the final endpoint; this may be assessed using observational studies or arm-based analyses of trials which have measured both the surrogate and the final outcome. This level of validation requires an assessment of the individual-level (absolute) association between endpoints, and is usually undertaken using correlation analysis. Level 1 of the hierarchy represents the strongest level of surrogate validation: in order to achieve this level of validation, the treatment effect on the surrogate must correspond to the treatment effect on the final outcome. Demonstrating this level of validity requires an analysis of correlation in terms of treatment effects between arms based on data from RCTs (trial-level association). Other validation frameworks have been proposed to assess the strength of association between surrogate and final endpoints. These include the criteria proposed by the German Institute of Quality and Efficiency in Health Care[10] (IQWiG; based on the treatment effect association only) and the Biomarker-Surrogate Evaluation Schema criteria[11] (BSES2; based on both absolute and treatment effect associations). These frameworks differ in terms of the types of analyses and the strength of the relationship required to determine the reliability of the surrogate.

This systematic review summarises published meta-regression studies reporting correlation and regression analyses for the strength of the association between response-based outcomes and PFS, TTP

or OS in (primarily) advanced or metastatic cancer, across any tumour site, in order to assess whether response-based outcomes may be considered as valid surrogates for PFS, TTP or OS.

Methods

Inclusion and exclusion criteria

Inclusion was restricted to articles reporting meta-analyses or meta-regressions across multiple studies, and reporting the strength of association between response outcomes (ORR, CR, PR, VGPR or DoR) and either PFS, TTP or OS. The included meta-regressions could themselves include RCTs and/or single-arm studies. However, individual reports analysing single trials or single cohorts were excluded from this review. Included meta-analyses could report absolute associations and/or treatment effect associations. These associations had to be reported as a correlation coefficient (e.g. Pearson r or Spearman r_s) and/or a coefficient of determination (R^2) between relevant outcomes.

Studies of any cancer and any treatment were included. The review focussed mainly on studies of advanced or metastatic cancers (and/or treatment with palliative intent), as these studies were more likely to report PFS and OS. However, studies reporting relevant outcomes were included even where the stage was not specifically restricted to advanced/metastatic disease for all patients or where this was unclear (this applied particularly to haematological cancers). Studies were excluded if they explicitly referred to adjuvant or neo-adjuvant treatment, or treatments which are given with curative intent. Studies were only included if they were written on English or contained sufficient detail in English.

The review protocol is registered on PROSPERO with registration number CRD42019127606.

Search strategy

Five databases (MEDLINE, EMBASE, Web of Science, the Cochrane Database of Systematic Reviews and CINAHL) were searched from inception to March 2019. Search terms included: cancer terms AND response terms AND terms for PFS, TTP and/or OS AND terms for regression, correlation, prediction,

association or relationship AND terms for endpoint and/or surrogate. Search results were limited to the English language and to studies undertaken in humans. The MEDLINE search strategy is provided in Supplementary Information 1. In addition, a citation search was undertaken based on two existing meta-reviews of surrogate relationships; this identified studies which have cited any of the 48 articles included in the review by Fischer *et al*. (2016)[12] and/or any of the 19 articles included in the review by Davis *et al*. (2012).[13] In addition, relevant existing meta-reviews, including Fischer *et al.* (2016),[12] Davis *et al*. (2012),[13] Savina *et al*. (2018),[14] Haslam *et al*. (2019)[15] and any reviews identified during searching, were checked for relevant studies.

Scoring the strength of association: IQWiG and BSES2 scoring

In this review, two sets of published criteria were used to assess the strength of association between surrogate and final endpoints: the IQWiG criteria[10] and the BSES2 criteria.[11]

The IQWiG criteria[10] are based on the correlation coefficient (r) for the treatment effect association. Where r was not reported, it was calculated as the square-root of R^2, if available. As the Medium score bracket was not clearly defined, slight modifications were made to the IQWiG criteria based on the approach used in the previous review by Savina *et al*.[14] (Supplementary Table 1). The IQWiG score was generated based on the magnitude of r, irrespective of its sign (i.e. a negative correlation could generate a high score). The IQWiG criteria were scored as follows: High (lower confidence interval of r is ≥ 0.85); Medium+ ($r \geq 0.85$ with no reported confidence interval or $r \geq 0.85$ with wide confidence intervals [lower limit <0.85]); Medium ($0.85 > r \geq 0.7$ and upper confidence interval of r is ≥ 0.7 and lower confidence interval of r is < 0.85, or $0.85 > r \geq 0.7$ with no reported confidence interval); or Low (upper confidence interval of r is < 0.7 or $r < 0.7$ with no reported confidence interval).

The BSES2 criteria[11] require R^2 values for both the absolute and treatment effect associations. Where R^2 was not reported, it was calculated as the square of r, if available. BSES2 criteria were used as an adaptation from the original BSES criteria, as described in Savina *et al*. (2018).[14] The original BSES criteria require R^2 for both individual and treatment effect associations and a value for the surrogate threshold effect (STE). Since so few articles report STE, this review used BSES2, which does not

require the STE. The BSES2 criteria were scored as follows: Excellent (R^2 [treatment effect] ≥ 0.6 and R^2 [absolute] ≥ 0.6); Good (R^2 [treatment effect] ≥ 0.4 and R^2 [absolute] ≥ 0.4); Fair (R^2 [treatment effect] ≥ 0.2 and R^2 [absolute] ≥ 0.2); Poor (R^2 [treatment effect] < 0.2 and/or R^2 [absolute] < 0.2). Further details on the IQWiG and BSES2 scoring systems are provided in Supplementary Tables 1 and 2.

Study selection and data extraction

Titles and abstracts of articles retrieved by the search were examined by one reviewer and a subset were checked by a second reviewer early in the process, followed by a discussion to ensure consistency in the selection decisions. Full texts were examined by one reviewer and a subset were checked by a second reviewer, with any discrepancies resolved through discussion.

Data were extracted by one reviewer and all data were checked by a second reviewer. Data were extracted relating to study design, participant characteristics, surrogate and final endpoints analysed, methods for correlation and regression, and results including absolute associations, associations between treatment effects, STE and regression equations.

Data synthesis

Data were presented in a narrative synthesis. Plots were constructed to illustrate the reported associations within each study. Some of the included meta-regression studies reported multiple subgroup analyses with differing results. Therefore, each horizontal row in the plots illustrates the range of reported associations across all subgroup analyses within a single meta-regression study. Where an included meta-regression study reported on more than one cancer type, these are shown on separate rows on the plots.

For associations between absolute values of endpoints, the plots show the range of correlation coefficients per study, across all subgroup analyses. All types of correlation coefficient were included,

e.g. Pearson r and Spearman r_s. If no correlation coefficient was reported, then Pearson r was calculated as the square-root of R^2, if available.

For associations between treatment effects, the plots show the range of regression coefficients of determination (R^2) per study, across all subgroup analyses. The plots include both adjusted and unadjusted R^2 values, as well as values from weighted and unweighted regressions. For studies in which R^2 was not reported, this was calculated as the square of the Pearson r correlation coefficient, if available. R^2 was not calculated from other correlation coefficients such as Spearman, or where the method of correlation was unclear.

Quality assessment

Included meta-regression studies were assessed for methodological quality based on key criteria from the AMSTAR-2[16] and ReSEEM[17] checklists most relevant to our review.

Results

Number of included meta-regression studies

The literature search generated 2,829 citations (Figure 1), of which 2,630 were excluded during the review of titles and abstracts and a further 135 excluded during the review of full texts. In total, 63 studies (within 64 references) were included in the review.[18-81]

Characteristics of included meta-regression studies

Summaries of study characteristics and reported data types are provided in Supplementary Tables 3 and 4 respectively, while full details of study characteristics for each of the 63 included studies are provided in Supplementary Table 5.

The most commonly reported surrogate relationships were ORR to OS (57 studies), ORR to PFS (22 studies), CR to OS (8 studies) and CR to PFS (7 studies). Other response outcomes (DoR, PR, VGPR/CR) were only reported in 1-2 studies each. Twenty different cancer types were analysed, the most common being NSCLC (16 studies), CRC (10 studies), various solid tumours (8 studies) and

breast cancer (5 studies). Disease stage was advanced/metastatic in 43 studies and unclear in 9 studies, while the remainder (11 studies) gave other descriptions mostly indicating advanced, extensive or recurrent disease. Treatment was first-line in 23 studies, later lines or combinations of lines in 32 studies, and not reported in 8 studies. Treatment type was chemotherapy in 21 studies, immune checkpoint inhibitors in 9 studies, targeted therapy in 8 studies, and various other treatment combinations in the remainder.

The various meta-regressions included between 4 and 191 primary studies and between 407 and 44,125 patients each. The majority of meta-regressions (N=44) included only RCTs, while 17 included both RCTs and single-arm studies and 2 included only single-arm studies. Most meta-regressions (N=58) analysed aggregate data (e.g. medians or other summary measures per study arm), whilst 5 analysed individual patient data (IPD). Across all meta-regressions, 32 reported absolute (individual-level) associations, 38 reported treatment effect (trial-level) associations, and only 4 reported the STE.

Methodological quality of included meta-regression studies

Methodological quality of the included studies is shown in Supplementary Table 6. All studies had clear inclusion criteria; 65% reported a comprehensive literature search; and 98% reported a correlation coefficient or R^2 value (the one study not reporting these was included as it reported a regression slope). However, only 27% reported duplicate study selection; 48% reported duplicate data extraction or checking; and 13% reported a risk of bias assessment of included studies. In addition, only 37% explored heterogeneity through subgroup analyses, and only 40% reported confidence intervals around the correlation coefficient or R^2.

Results of included studies

The reported associations between surrogate and final endpoints are summarised in Table 1 and illustrated in Figure 2 to Figure 5. Full results for each included meta-regression study are provided in Supplementary Table 7 (for absolute associations) and Supplementary Table 8 (for treatment effect associations).

Absolute (individual-level) correlation and regression

The range of absolute (individual-level) correlation coefficients is summarised in Table 1 and illustrated in Figure 2 (for the association between ORR and PFS) and Figure 3 (for the association between ORR and OS). Some of the included meta-regression studies reported multiple subgroup analyses with differing results. Therefore, each horizontal row in the plots illustrates the range of correlation coefficients across all subgroup analyses within a single meta-regression study. Where an included meta-regression reported on more than one cancer type, these are shown on separate rows on the plots. It is worth noting that the included meta-regression studies differed in terms of various factors, such as the number of included primary studies (shown as N on the plots), treatment type, line of treatment and precise clinical population (full details in Supplementary Table 7).

ORR and PFS (or TTP): The reported correlation coefficients (Pearson r or Spearman r_s) between absolute ORR and PFS ranged from -0.72 to 0.96, based on multiple analyses within 12 studies across 10 cancer types[44,45,52,54,55,59,62,63,65,66,72,78] (Figure 2 and Table 1). Across those studies which report only a single analysis, the correlation coefficient was generally above 0.60; however, some estimates were lower. Confidence intervals around the correlation coefficients were rarely reported. Few separate meta-regressions reported on the same tumour site, hence it is difficult to assess whether ORR may be a more reliable surrogate in certain cancer types than others. One study reported on ORR and TTP (gastric cancer; correlation r_s = 0.41 to 0.56 across subgroup analyses, not shown on the plot).[42]

ORR and OS: The reported correlation coefficients between absolute ORR and OS ranged from -0.40 to 1.00, based on 27 studies across 15 cancer types[18,43,19,20,35,37,38,42,45,49-52,59-66,68,70-72,75,78] (Figure 3 and Table 1). Confidence intervals around the correlation coefficients, where reported, were generally fairly

238 wide. The majority of correlation coefficients were above 0.40; however, several estimates were lower.
239 Neither the correlation coefficients reported from multiple analyses within the same study, nor those
240 reported across separate studies, suggested a clear pattern by cancer type.

241 **CR and PFS or OS:** The correlation coefficients between absolute CR and PFS in two studies of small-
242 cell lung cancer (SCLC)[59] and non-Hodgkin's lymphoma (NHL)[81] ranged from 0.22 to 0.83, while the
243 correlation coefficients between absolute CR and OS ranged from -0.04 to 0.62, based on 3 studies of
244 NSCLC,[49] SCLC[59] and gastroesophageal cancer[61] (Table 1).

245 **PR and PFS or OS:** The correlation coefficient between absolute PR and PFS ranged from 0.35 to 0.70
246 across subgroup analyses within one study of SCLC,[59] while the highest correlation coefficient between
247 absolute PR and OS ranged from 0.29 to 0.66 in the same study[59] (Table 1).

248 **DoR and PFS or OS:** No studies reported on the absolute association between DoR and PFS or OS.

249 **Treatment effect (trial-level) correlation and regression**

250 The range of treatment effect (trial-level) R^2 values is summarised in Table 1 and illustrated in Figure
251 4Figure **4** (for the association between ORR and PFS) and Figure 5 (for the association between ORR
252 and OS). Each horizontal row in the plots illustrates the range of R^2 values across all subgroup analyses
253 within a single meta-regression study. Where an included meta-regression reported on more than one
254 cancer type, these are shown separately on the plots. It is worth noting that the meta-regressions differed
255 in terms of the number of included primary studies (shown as N on the plots), treatment type, line of
256 treatment and precise clinical population (full details in Supplementary Table 8).

257 **ORR and PFS:** The regression R^2 values for the treatment effect association between ORR and PFS
258 ranged from 0.18 to 0.94, based on 9 studies across 4 cancer types: NSCLC,[21,22,45,67,77] ovarian
259 cancer,[27,72] colorectal cancer[26,77] and various solid tumours[67,77,79] (Figure 4 and Table 1). The majority
260 of R^2 values were above 0.40. The R^2 values reported from multiple analyses within the same study,
261 and those reported across separate studies, did not suggest a clear pattern by cancer type. Confidence
262 intervals around the R^2 values, where reported, were generally fairly wide.

ORR and OS: The regression R^2 values for the treatment effect association between ORR and OS ranged from -0.08 to 0.84, based on 31 studies across 11 cancer types[21-23,25-32,34,36,37,39-41,45-47,53,56-58,60,63,67,73,74,77,79] (Figure 5 and Table 1). With the exception of one analysis,[77] all R^2 values were below 0.60. The R^2 values reported from multiple analyses within the same study, and those reported across separate studies, did not suggest a clear pattern by cancer type. Confidence intervals around the R^2 values, where reported, were generally wide.

CR and PFS or OS: The regression R^2 for the treatment effect association between CR and PFS ranged from 0.45 to 0.93 across subgroup analyses within one study of NHL,[69] while the regression R^2 for the treatment effect association between CR and OS within two studies of breast cancer[36] and SCLC[34] ranged from 0.05 to 0.48 (Table 1).

PR and PFS or OS: No studies reported the treatment effect association between PR and PFS or OS.

DoR and PFS or OS: No studies reported R^2 between DoR and OS or PFS. Two studies in colorectal cancer[29] and pancreatic cancer[28] reported Spearman correlation coefficients between DoR and OS ranging from 0.40 to 0.76 (Table 1).

Influence of clinical and study factors on association

The impact of the following patient and study factors on the association between ORR and OS was explored: treatment line; treatment type; response criteria; adjustment of OS for crossover and post-progression treatments; and aggregate versus IPD data (Supplementary Table 9). No clear effect on the association between ORR and OS was identified for any individual factor. However, this analysis was limited by the small number of publications assessing each factor within each cancer, and the wide ranges of associations observed for each.

Five of the 63 included meta-analyses analysed IPD rather than aggregate data; two in breast cancer (Bruzzi 2005[23] Burzykowski 2008[24]), one in colorectal cancer (Buyse 2000[25]), one in NHL (Shi 2017[69]) and one in ovarian cancer (Rose 2010[66]). The associations reported in these studies were not noticeably different to those in other studies (see Figures 2 to 5).

Regression equations

Regression equations were reported in fourteen studies for the relationship between ORR and OS; of these, four reported absolute associations[42,52,72,76] and ten reported treatment effect associations.[31-33,36,41,46,56,58,67,77] Regression equations were also reported in eight studies for the relationship between ORR and PFS; of these, four reported absolute associations[52,54,72,76] and four reported treatment effect associations.[24,33,67,77] These analyses spanned 10 cancer types. Full details are provided in Supplementary Tables 10 and 11. There was substantial variation in the effect measures used for both the surrogate and final outcomes; for example, difference in medians, hazard ratio (HR), odds ratio (OR), log-transformed or not. None of the included studies attempted to externally validate their regression equations for the relationship between response and other outcomes.

Surrogate threshold effect (STE)

The STE - the smallest treatment effect on the surrogate that predicts a non-zero treatment effect on the true endpoint[82] - was reported in only four studies (Supplementary Table 12).[26,39,69,77] For the relationship between ORR and PFS, one study[77] in various solid tumours reported that a difference in ORR of 15% would be required to predict a non-zero treatment effect on the HR for PFS. For the relationship between ORR and OS, two studies in various solid tumours[77] and NSCLC[39] reported that a difference in ORR of 21% and 55% respectively would be required to predict a non-zero treatment effect on the HR for OS, while one study[39] also reported that a difference in ORR of 41% would be required to predict a non-zero treatment effect on the difference in median OS. A further study in colorectal cancer[26] reported that an OR for ORR of 0.28 would be required to predict a non-zero treatment effect on the OR for OS. Finally, for the relationship between CR and PFS, one study in NHL[69] reported that an OR for CR (at 30 months) of 1.56 would be required to predict a non-zero treatment effect on the HR for PFS.

IQWiG and BSES2 scores for strength of association

IQWiG and BSES2 scores for the strength of association between surrogate and final endpoints were calculated for all reported subgroup analyses with sufficient data; therefore, meta-regression studies that reported more subgroups are more strongly represented in this analysis. These data are presented graphically in Supplementary Figures 1 and 2.

In terms of IQWiG scores, of 202 analyses (across 63 studies), 0 (0%) scored high, 15 (7%) scored medium+, 26 (13%) scored medium, 76 (38%) scored low and 85 (42%) were not evaluable. In terms of BSES2 scores, of 202 analyses (across 63 studies), 0 (0%) scored excellent, 3 (1%) scored good, 3 (1%) scored fair, 7 (3%) scored poor and 189 (94%) were not evaluable.

Discussion

This systematic review summarises published meta-regression studies reporting correlation and regression analyses for the strength of the association between response outcomes and PFS, TTP or OS across different types of cancer. In total, the review included 63 studies across 20 cancer types. The most commonly analysed relationships were between ORR and either PFS or OS.

For the association between ORR and PFS, the majority of reported correlation coefficients between absolute values were above 0.60 (range -0.72 to 0.96). For association between treatment effects on ORR and PFS, the majority of regression R^2 values were above 0.40 (range 0.18 to 0.94). The association between ORR and OS appeared weaker than that between ORR and PFS; while the majority of reported correlation coefficients between absolute values were above 0.40, several estimates were lower (range -0.40 to 1.00). For association between treatment effects on ORR and OS, all regression R^2 values except one were below 0.60 (range -0.08 to 0.84).

There was no clear pattern by cancer type for either the absolute or treatment effect associations, based on both multiple analyses within the same study and results across separate studies. Confidence

intervals around the reported correlation coefficients and R^2 values were generally wide and often not reported.

Strength of association across all subgroup analyses within all included meta-regression studies was assessed using the IQWiG and BSES2 scoring systems. The majority of analyses were not evaluable due to lack of required data. Of those analyses that could be scored, scores were relatively low, suggesting that response-based endpoints may be poor surrogates for OS.

Previous systematic reviews of surrogate endpoints in advanced cancer have been published. Savina *et al* and Haslam *et al* have reported systematic reviews of meta-analyses assessing any endpoint as a surrogate for OS.[14,15] Both of these reviews also assessed the strength of association using surrogate validation frameworks; both studies used adaptations of the IQwiG framework, and Savina *et al* also used the BSES2 framework. These previous reviews generally focussed on the main analyses presented within individual meta-analyses (usually that with the largest number of patients). Similar to our review, these previous reviews suggested that response-based outcomes are likely to be poor surrogates for OS. Our systematic review focusses exclusively on response-based surrogates; it includes a comprehensive search to identify relevant studies, considers PFS as a potential final endpoint as well as OS, is more up to date, includes a greater number of studies, and reports results for the full breadth of analyses reported in the included meta-regression studies compared with these previous reviews. This provides a more complete picture of the extent of heterogeneity in reported relationships across the full range of meta-analyses across each cancer area. This additional breadth provides a better basis to inform judgements about the validity of response-based endpoints as a surrogate for PFS or OS.

The review is subject to a number of limitations. The reported data were highly heterogeneous in terms of effect measure and method of analysis. Therefore, some simplifying assumptions had to be made to allow the data to be summarised. For example, correlation coefficients were summarised regardless of method (Pearson, Spearman or other); R^2 values were summarised irrespective of whether or not the regression was weighted and whether or not the R^2 was adjusted; and for treatment effect associations, R^2 values were summarised regardless of the effect measure (e.g. HR, OR, difference in medians). In

addition, only five studies used IPD rather than aggregate data in their analysis; this is a limitation of the analyses conducted in the majority of meta-reviews. A recent review by Xie *et al* highlighted wide variability in reporting standards across surrogate evaluation meta-regression studies; future analyses should attempt to adhere to current best practice, for example, the reporting of surrogate endpoint evaluation using meta-analyses (ReSEEM) guidelines in order to improve the quality of these analyses.[17]

It should further be noted that whilst meta-regression has been widely used for the purpose of evaluating the validity of surrogate endpoints in oncology, this method has been criticised as it ignores uncertainty around the treatment effect on the surrogate outcome (which is treated as a fixed covariate in the analysis). Newer methods, such as the bivariate random effects meta-analysis (BRMA) model reported by Bujkiewicz *et al*,[83] provides an approach for both the validation and prediction of surrogate endpoints within a Bayesian framework. This approach allows for borrowing of information across studies and fully accounts for all uncertainty surrounding the surrogate relationship. In spite of the generally poor association between response-based outcomes and final outcomes, there may still be instances in which generating predictions on the basis of response is necessary; for example, within health economic models, or more broadly, for decision-making within health technology assessment. In instances where the surrogate association is weak, this uncertainty would manifest as a wider prediction interval. If such predictions are necessary, it is therefore important that all uncertainty is reflected in the model. Future surrogate evaluation studies should consider the use of the BRMA model, rather than conventional meta-regression, as a means of fully reflecting this uncertainty.

Conclusions

This systematic review suggests that response-based endpoints such as ORR and CR may not be reliable surrogates for PFS or OS in cancer treatment. Strength of association varied widely between and within studies, with no clear pattern by cancer type. The strength of association between ORR and OS appeared weaker and more variable than that between ORR and PFS, both for associations between absolute

389 endpoints and associations between treatment effects. Whilst there may still be value in using response
390 outcomes as a means of predicting final outcomes such as OS, it is important that those predictions are
391 made on the basis of models which fully reflect the uncertainty around the treatment effect on the
392 surrogate outcome.

393

394 ## Additional Information

395 **Acknowledgements**

396 Thanks to Andrea Shippam for administrative and Endnote support.

397 **Authors' contributions**

398 KC and PT designed the protocol, selected studies, and extracted and analysed data. AC designed and
399 undertook the literature searches. KE selected studies and extracted data. All authors inputted to the
400 manuscript.

401 **Ethics approval and consent to participate**

402 Ethics approval was not required since all data were already in the public domain.

403 **Consent for publication**

404 Consent for publication was not required since all data were already in the public domain.

405 **Data availability**

406 All data are provided in the tables, figures and supplementary information.

407 **Conflict of interest**

408 The authors declare no conflicts of interest.

409 **Funding**

410 This study was funded by the UK National Institute for Health Research (NIHR) Health Technology
411 Assessment (HTA) Programme, project number NIHR127852.

412

413 ## References

414 1. Fleming TR, Powers JH. Biomarkers and surrogate endpoints in clinical trials. *Statistics in*
415 *Medicine*; **31**(25), 2973-2984 (2012).

416 2. Taylor RS, Elston J. The use of surrogate outcomes in model-based cost-effectiveness
417 analyses: a survey of UK Health Technology Assessment reports. *Health Technology*
418 *Assessment*; **13**(8), 1-72 (2009).

419 3. Fleming TR, DeMets DL. Surrogate end points in clinical trials: are we being misled? *Annals*
420 *of Internal Medicine*; **125**(7), 605-613 (1996).

421 4. Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria.
422 *Statistics in Medicine*; **8**(4), 431-440 (1989).

423 5. Buyse M, Molenberghs G, Burzykowski T, Renard D, Geys H. The validation of surrogate
424 endpoints in meta-analyses of randomized experiments. *Biostatistics*; **1**(1), 49-67 (2000).

425 6. Heller G. Statistical controversies in clinical research: an initial evaluation of a surrogate end
426 point using a single randomized clinical trial and the Prentice criteria. *Annals of Oncology*;
427 **26**(10), 2012-2016 (2015).

428 7. Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: preferred
429 definitions and conceptual framework. *Clinical Pharmacology & Therapeutics*; **69**(3), 89-95
430 (2001).

431 8. Bucher HC, Guyatt GH, Cook DJ, Holbrook A, McAlister FA. Users' guides to the medical
432 literature: XIX. Applying clinical trial results. A. How to use an article measuring the effect
433 of an intervention on surrogate end points. *JAMA*; **282**(8), 771-778 (1999).

434 9. Elston J, Taylor RS. Use of surrogate outcomes in cost-effectiveness models: a review of
435 United Kingdom health technology assessment reports. *Int J Technol Assess Health Care*;
436 **25**(1), 6-13 (2009).

437 10. German Institute of Quality and Efficiency in Health Care (IQWiG). Validity of surrogate
438 endpoints in oncology. (2011).

439 11. Lassere MN, Johnson KR, Schiff M, Rees D. Is blood pressure reduction a valid surrogate
440 endpoint for stroke prevention? An analysis incorporating a systematic review of randomised
441 controlled trials, a by-trial weighted errors-in-variables regression, the surrogate threshold
442 effect (STE) and the Biomarker-Surrogacy (BioSurrogate) Evaluation Schema (BSES). *BMC*
443 *Med Res Methodol*; **12**, 27 (2012).

444 12. Fischer A, Hernandez-Villafuerte K, Latimer N, Henshall C. Extrapolation from Progression-
445 Free Survival to Overall Survival in Oncology. Office of Health Economics;(2016).

446 13. Davis S, Tappenden P, Cantrell A. A review of studies examining the relationship between
447 progression-free survival and overall survival in advanced or metastatic cancer. University of
448 Sheffield report for NICE Decision Support Unit (2012).

449 14. Savina M, Gourgou S, Italiano A, Dinart D, Rondeau V, Penel N *et al.* Meta-analyses
450 evaluating surrogate endpoints for overall survival in cancer randomized trials: A critical
451 review. *Critical Reviews in Oncology Hematology*; **123**, 21-41 (2018).

452 15. Haslam A, Hey SP, Gill J, Prasad V. A systematic review of trial-level meta-analyses
453 measuring the strength of association between surrogate end-points and overall survival in
454 oncology. *European Journal of Cancer*; **106**, 196-211 (2019).

16. Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J *et al.* AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ*; **358**, j4008 (2017).

17. Xie W, Halabi S, Tierney JF, Sydes MR, Collette L, Dignam JJ *et al.* A Systematic Review and Recommendation for Reporting of Surrogate Endpoint Evaluation using Meta-analyses (ReSEEM). *JNCI Cancer Spectrum* (2019).

18. Abdel-Rahman O. Surrogate end points for overall survival in trials of PD-(L)1 inhibitors for urinary cancers: a systematic review. *Immunotherapy*; **10**(2), 139-148 (2018).

19. Agarwal N, Bellmunt J, Maughan BL, Boucher KM, Choueiri TK, Qu AQ *et al.* Six-month progression-free survival as the primary endpoint to evaluate the activity of new agents as second-line therapy for advanced urothelial carcinoma. *Clinical Genitourinary Cancer*; **12**(2), 130-137 (2014).

20. Agarwal SK, Mangal N, Menon RM, Freise KJ, Salem AH. Response Rates as Predictors of Overall Survival: A Meta-Analysis of Acute Myeloid Leukemia Trials. *J Cancer*; **8**(9), 1562-1567 (2017).

21. Blumenthal GM, Karuri SW, Zhang H, Zhang LJ, Khozin S, Kazandjian D *et al.* Overall Response Rate, Progression-Free Survival, and Overall Survival With Targeted and Standard Therapies in Advanced Non-Small-Cell Lung Cancer: US Food and Drug Administration Trial-Level and Patient-Level Analyses. *Journal of Clinical Oncology*; **33**(9), 1008 (2015).

22. Blumenthal GM, Zhang L, Zhang H, Kazandjian D, Khozin S, Tang S *et al.* Milestone Analyses of Immune Checkpoint Inhibitors, Targeted Therapy, and Conventional Therapy in Metastatic Non-Small Cell Lung Cancer Trials: A Meta-analysis. *JAMA Oncology*; **3**(8), e171029 (2017).

23. Bruzzi P, Del Mastro L, Sormani MP, Bastholt L, Danova M, Focan C *et al.* Objective response to chemotherapy as a potential surrogate end point of survival in metastatic breast cancer patients. *Journal of Clinical Oncology*; **23**(22), 5117-5125 (2005).

24. Burzykowski T, Buyse M, Piccart-Gebhart MJ, Sledge G, Carmichael J, Luck HJ *et al.* Evaluation of tumor response, disease control, progression-free survival, and time to progression as potential surrogate end points in metastatic breast cancer. *Journal of Clinical Oncology*; **26**(12), 1987-1992 (2008).

25. Buyse M, Thirion P, Carlson RW, Burzykowski T, Molenberghs G, Piedbois P. Relation between tumour response to first-line chemotherapy and survival in advanced colorectal cancer: A meta-analysis. *Lancet*; **356**(9227), 373-378 (2000).

26. Ciani O, Buyse M, Garside R, Peters J, Saad ED, Stein K *et al.* Meta-analyses of randomized controlled trials show suboptimal validity of surrogate outcomes for overall survival in advanced colorectal cancer. *Journal of Clinical Epidemiology*; **68**(7), 833-842 (2015).

27. Colloca G, Venturino A. Trial-level analysis of progression-free survival and response rate as end points of trials of first-line chemotherapy in advanced ovarian cancer. *Medical Oncology*; **34**(5), 87 (2017).

28. Colloca G, Venturino A, Guarneri D. Analysis of Response-Related and Time-to-event Endpoints in Randomized Trials of Gemcitabine-Based Treatment Versus Gemcitabine Alone as First-Line Treatment of Patients With Advanced Pancreatic Cancer. *Clinical Colorectal Cancer*; **15**(3), 264-276 (2016a).

29. Colloca G, Venturino A, Guarneri D. Analysis of Clinical End Points of Randomised Trials Including Bevacizumab and Chemotherapy versus Chemotherapy as First-line Treatment of Metastatic Colorectal Cancer. *Clinical Oncology (Royal College of Radiologists)*; **28**(10), e155-164 (2016b).

30. Colloca G, Vitucci P, Venturino A. Trial Level Analysis of Prostate-Specific Antigen-Related Versus Unrelated Endpoints in Phase III Trials of First-Line and Second-Line Medical Treatments of Patients With Metastatic Castration-Resistant Prostate Cancer. *Clinical Genitourinary Cancer*; **14**(5), 389-397 (2016c).

31. Cremolini C, Antoniotti C, Pietrantonio F, Berenato R, Tampellini M, Baratelli C *et al.* Surrogate Endpoints in Second-Line Trials of Targeted Agents in Metastatic Colorectal Cancer: A Literature-Based Systematic Review and Meta-Analysis. *Cancer Research & Treatment*; **49**(3), 834-845 (2017).

32. Delea TE, Khuu A, Heng DYC, Haas T, Soulieres D. Association between treatment effects on disease progression end points and overall survival in clinical studies of patients with metastatic renal cell carcinoma. *British Journal of Cancer*; **107**(7), 1059-1068 (2012).

33. Elia EG, Städler N, Ciani O, Taylor RS, Bujkiewicz S. Combining tumour response and progression free survival as surrogate endpoints for overall survival in advanced colorectal cancer. *arXiv preprint arXiv:180902935* (2018).

34. Foster NR, Qi Y, Shi Q, Krook JE, Kugler JW, Jett JR *et al.* Tumor response and progression-free survival as potential surrogate endpoints for overall survival in extensive stage small-cell lung cancer: findings on the basis of North Central Cancer Treatment Group trials. *Cancer*; **117**(6), 1262-1271 (2011).

35. Giessen C, Laubender RP, Ankerst DP, Stintzing S, Modest DP, Schulz C *et al.* Surrogate endpoints in second-line treatment for mCRC: a systematic literature-based analysis from 23 randomised trials. *Acta Oncologica*; **54**(2), 187-193 (2015).

36. Hackshaw A, Knight A, Barrett-Lee P, Leonard R. Surrogate markers and survival in women receiving first-line combination anthracycline chemotherapy for advanced breast cancer. *British Journal of Cancer*; **93**(11), 1215-1221 (2005).

37. Hamada T, Nakai Y, Isayama H, Yasunaga H, Matsui H, Takahara N *et al.* Progression-free survival as a surrogate for overall survival in first-line chemotherapy for advanced pancreatic cancer. *European Journal of Cancer*; **65**, 11-20 (2016).

38. Han K, Ren M, Wick W, Abrey L, Das A, Jin J *et al.* Progression-free survival as a surrogate endpoint for overall survival in glioblastoma: a literature-based meta-analysis from 91 trials. *Neuro-Oncology*; **16**(5), 696-706 (2014).

39. Hashim M, Pfeiffer BM, Bartsch R, Postma M, Heeg B. Do Surrogate Endpoints Better Correlate with Overall Survival in Studies That Did Not Allow for Crossover or Reported Balanced Postprogression Treatments? An Application in Advanced Non-Small Cell Lung Cancer. *Value in Health*; **21**(1), 9-17 (2018).

40. Hotta K, Kato Y, Leighl N, Takigawa N, Gaafar RM, Kayatani H *et al.* Magnitude of the Benefit of Progression-Free Survival as a Potential Surrogate Marker in Phase 3 Trials Assessing Targeted Agents in Molecularly Selected Patients with Advanced Non-Small Cell Lung Cancer: Systematic Review. *Plos One*; **10**(3) (2015).

41. Hotta K, Kiura K, Fujiwara Y, Takigawa N, Oze I, Ochi N *et al.* Association between incremental gains in the objective response rate and survival improvement in phase III trials of first-line chemotherapy for extensive disease small-cell lung cancer. *Annals of Oncology*; **20**(5), 829-834 (2009).

42. Ichikawa W, Sasaki Y. Correlation between tumor response to first-line chemotherapy and prognosis in advanced gastric cancer patients. *Annals of Oncology*; **17**(11), 1665-1672 (2006).

43. Imaoka H, Sasaki M, Takahashi H, Hashimoto Y, Ohno I, Mitsunaga S *et al.* Progression-free survival as a surrogate endpoint in advanced neuroendocrine neoplasms. *Endocrine-Related Cancer*; **24**(9), 475-483 (2017).

44. Imaoka H, Sasaki M, Takahashi H, Hashimoto Y, Ohno I, Mitsunaga S *et al.* Alternate Endpoints for Phase II Trials in Advanced Neuroendocrine Tumors. *Oncologist*; **24**(1), 47-53 (2019).

45. Ito K, Miura S, Sakaguchi T, Murotani K, Horita N, Akamatsu H *et al.* The impact of high PD-L1 expression on the surrogate endpoints and clinical outcomes of anti-PD-1/PD-L1 antibodies in non-small cell lung cancer. *Lung Cancer*; **128**, 113-119 (2019).

46. Johnson KR, Ringland C, Stokes BJ, Anthony DM, Freemantle N, Irs A *et al.* Response rate or time to progression as predictors of survival in trials of metastatic colorectal cancer or non-small-cell lung cancer: a meta-analysis. *Lancet Oncology*; **7**(9), 741-746 (2006).

47. Kaufman HL, Schwartz LH, William WN, Jr., Sznol M, Fahrbach K, Xu Y *et al.* Evaluation of classical clinical endpoints as surrogates for overall survival in patients treated with immune checkpoint blockers: a systematic review and meta-analysis. *Journal of Cancer Research & Clinical Oncology*; **144**(11), 2245-2261 (2018).

48. Lee L, Wang L, Crump M. Identification of potential surrogate end points in randomized clinical trials of aggressive and indolent non-Hodgkin's lymphoma: correlation of complete response, time-to-event and overall survival end points. *Annals of Oncology*; **22**(6), 1392-1403 (2011).

49. Li J, He Q, Yu X, Khan K, Weng XW, Guan MJ. Complete response associated with immune checkpoint inhibitors in advanced non-small-cell lung cancer: a meta-analysis of nine randomized controlled trials. *Cancer Management and Research*; **11**, 1623-1629 (2019).

50. Li X, Liu S, Gu H, Wang D. Surrogate end points for survival in the target treatment of advanced non-small-cell lung cancer with gefitinib or erlotinib. *Journal of Cancer Research & Clinical Oncology*; **138**(11), 1963-1969 (2012).

51. Liu L, Chen F, Zhao J, Yu H. Correlation between overall survival and other endpoints in metastatic breast cancer with second- or third-line chemotherapy: Literature-based analysis of 24 randomized trials. *Bulletin du Cancer*; **103**(4), 336-344 (2016).

52. Louvet C, de Gramont A, Tournigand C, Artru P, Maindrault-Goebel F, Krulik M. Correlation between progression free survival and response rate in patients with metastatic colorectal carcinoma. *Cancer*; **91**(11), 2033-2038 (2001).

53. Makris EA, MacBarb R, Harvey DJ, Poultsides GA. Surrogate End Points for Overall Survival in Metastatic, Locally Advanced, or Unresectable Pancreatic Cancer: A Systematic Review and Meta-Analysis of 24 Randomized Controlled Trials. *Annals of Surgical Oncology*; **24**(8), 2371-2378 (2017).

54. Mangal N, Salem AH, Li MY, Menon R, Freise KJ. Relationship between response rates and median progression-free survival in non-Hodgkin's lymphoma: A meta-analysis of published clinical trials. *Hematological Oncology*; **36**(1), 37-43 (2018).

55. Mangal N, Salem AH, Menon RM, Freise KJ. Use of depth of response to predict progression-free survival in relapsed or refractory multiple myeloma: Evaluation of results from 102 clinical trials. *Hematological Oncology*; **36**(3), 547-553 (2018).

56. Moriwaki T, Yamamoto Y, Gosho M, Kobayashi M, Sugaya A, Yamada T *et al.* Correlations of survival with progression-free survival, response rate, and disease control rate in advanced biliary tract cancer: a meta-analysis of randomised trials of first-line chemotherapy. *British Journal of Cancer*; **114**(8), 881-888 (2016).

57. Mushti SL, Mulkey F, Sridhara R. Evaluation of Overall Response Rate and Progression-Free Survival as Potential Surrogate Endpoints for Overall Survival in Immunotherapy Trials. *Clinical Cancer Research*; **24**(10), 2268-2275 (2018).

58. Nakashima K, Horita N, Nagai K, Manabe S, Murakami S, Ota E *et al.* Progression-Free Survival, Response Rate, and Disease Control Rate as Predictors of Overall Survival in Phase

III Randomized Controlled Trials Evaluating the First-Line Chemotherapy for Advanced, Locally Advanced, and Recurrent Non-Small Cell Lung Carcinoma. *Journal of Thoracic Oncology: Official Publication of the International Association for the Study of Lung Cancer*; **11**(9), 1574-1585 (2016).

59. Nickolich M, Babakoohi S, Fu P, Dowlati A. Clinical trial design in small cell lung cancer: surrogate end points and statistical evolution. *Clinical Lung Cancer*; **15**(3), 207-212 (2014).

60. Nie RC, Chen FP, Yuan SQ, Luo YS, Chen S, Chen YM *et al.* Evaluation of objective response, disease control and progression-free survival as surrogate end-points for overall survival in anti-programmed death-1 and anti-programmed death ligand 1 trials. *European Journal of Cancer*; **106**, 1-11 (2019).

61. Pang Y, Shen Z, Sun J, Wang W. Does the use of targeted agents in advanced gastroesophageal cancer increase complete response? A meta-analysis of 18 randomized controlled trials. *Cancer Management and Research*; **10**, 5505-5514 (2018).

62. Penel N, Ryckewaert T, Kramar A. What is an active regimen in carcinoma of unknown primary sites? Analysis of correlation between activity endpoints reported in phase II trials. Correlation of activity endpoints in phase II trials. *Bulletin du Cancer*; **101**(4), E19-24 (2014).

63. Petrelli F, Barni S. Surrogate end points and postprogression survival in renal cell carcinoma: an analysis of first-line trials with targeted therapies. *Clinical Genitourinary Cancer*; **11**(4), 385-389 (2013).

64. Petrelli F, Barni S. Surrogate endpoints in metastatic breast cancer treated with targeted therapies: an analysis of the first-line phase III trials. *Medical Oncology*; **31**(1), 776 (2014).

65. Ritchie G, Gasper H, Man J, Lord S, Marschner I, Friedlander M *et al.* Defining the most appropriate primary end point in phase 2 trials of immune checkpoint inhibitors for advanced solid cancers a systematic review and meta-analysis. *JAMA Oncology*; **4**(4), 522-528 (2018).

66. Rose PG, Tian CQ, Bookman MA. Assessment of tumor response as a surrogate endpoint of survival in recurrent/platinum-resistant ovarian carcinoma: A Gynecologic Oncology Group study. *Gynecologic Oncology*; **117**(2), 324-329 (2010).

67. Roviello G, Andre F, Venturini S, Pistilli B, Curigliano G, Cristofanilli M *et al.* Response rate as a potential surrogate for survival and efficacy in patients treated with novel immune checkpoint inhibitors: A meta-regression of randomised prospective studies. *European Journal of Cancer*; **86**, 257-265 (2017).

68. Sekine I, Tamura T, Kunitoh H, Kubota K, Shinkai T, Kamiya Y *et al.* Progressive disease rate as a surrogate endpoint of phase II trials for non-small-cell lung cancer. *Annals of Oncology*; **10**(6), 731-733 (1999).

69. Shi Q, Flowers CR, Hiddemann W, Marcus R, Herold M, Hagenbeek A *et al.* Thirty-Month Complete Response as a Surrogate End Point in First-Line Follicular Lymphoma Therapy: An Individual Patient-Level Analysis of Multiple Randomized Trials. *Journal of Clinical Oncology*; **35**(5), 552-560 (2017).

70. Shitara K, Matsuo K, Muro K, Doi T, Ohtsu A. Correlation between overall survival and other endpoints in clinical trials of second-line chemotherapy for patients with advanced gastric cancer. *Gastric Cancer*; **17**(2), 362-370 (2014).

71. Shukuya T, Mori K, Amann JM, Bertino EM, Otterson GA, Shields PG *et al.* Relationship between Overall Survival and Response or Progression-Free Survival in Advanced Non-Small Cell Lung Cancer Patients Treated with Anti-PD-1/PD-L1 Antibodies. *Journal of Thoracic Oncology: Official Publication of the International Association for the Study of Lung Cancer*; **11**(11), 1927-1939 (2016).

643 72. Siddiqui MK, Tyczynski J, Pahwa A, Fernandes AW. Objective response rate is a possible
644 surrogate endpoint for survival in patients with advanced, recurrent ovarian cancer.
645 *Gynecologic Oncology*; **146**(1), 44-51 (2017).

646 73. Sidhu R, Rong A, Dahlberg S. Evaluation of progression-free survival as a surrogate endpoint
647 for survival in chemotherapy and targeted agent metastatic colorectal cancer trials. *Clinical*
648 *Cancer Research*; **19**(5), 969-976 (2013).

649 74. Tanaka K, Kawano M, Iwasaki T, Itonaga I, Tsumura H. Surrogacy of intermediate endpoints
650 for overall survival in randomized controlled trials of first-line treatment for advanced soft
651 tissue sarcoma in the pre- and post-pazopanib era: A meta-analytic evaluation. *BMC Cancer*;
652 **19 (1) (no pagination)**(56) (2019).

653 75. Tang PA, Bentzen SM, Chen EX, Siu LL. Surrogate end points for median overall survival in
654 metastatic colorectal cancer: literature-based analysis from 39 randomized controlled trials of
655 first-line chemotherapy. *Journal of Clinical Oncology*; **25**(29), 4562-4568 (2007).

656 76. Tsujino K, Kawaguchi T, Kubo A, Aono N, Nakao K, Koh Y *et al.* Response Rate Is
657 Associated with Prolonged Survival in Patients with Advanced Non-small Cell Lung Cancer
658 Treated with Gefitinib or Erlotinib. *Journal of Thoracic Oncology*; **4**(8), 994-1001 (2009).

659 77. Tsujino K, Shiraishi J, Tsuji T, Kurata T, Kawaguchi T, Kubo A *et al.* Is response rate
660 increment obtained by molecular targeted agents related to survival benefit in the phase III
661 trials of advanced cancer? *Annals of Oncology*; **21**(8), 1668-1674 (2010).

662 78. Vidaurre T, Wilkerson J, Simon R, Bates SE, Fojo T. Stable disease is not preferentially
663 observed with targeted therapies and as currently defined has limited value in drug
664 development. *Cancer Journal*; **15**(5), 366-373 (2009).

665 79. Wilkerson J, Fojo T. Progression-free survival is simply a measure of a drug's effect while
666 administered and is not a surrogate for overall survival. *Cancer Journal*; **15**(5), 379-385
667 (2009).

668 80. Zer A, Prince RM, Amir E, Abdul Razak A. Evolution of Randomized Trials in
669 Advanced/Metastatic Soft Tissue Sarcoma: End Point Selection, Surrogacy, and Quality of
670 Reporting. *Journal of Clinical Oncology*; **34**(13), 1469-1475 (2016).

671 81. Zhu R, Lu D, Chu YW, Chai A, Green M, Zhang N *et al.* Assessment of Correlation Between
672 Early and Late Efficacy Endpoints to Identify Potential Surrogacy Relationships in Non-
673 Hodgkin Lymphoma: a Literature-Based Meta-analysis of 108 Phase II and Phase III Studies.
674 *Aaps Journal*; **19**(3), 669-681 (2017).

675 82. Buyse M, Molenberghs G, Paoletti X, Oba K, Alonso A, Van der Elst W *et al.* Statistical
676 evaluation of surrogate endpoints with examples from cancer clinical trials. *Biometrical*
677 *Journal*; **58**(1), 104-132 (2016).

678 83. Bujkiewicz S, Achana F, Papanikos T, Riley RD, Abrams KR. Technical Support Document
679 20: Multivariate meta-analysis of summary data for combining treatment effects on correlated
680 outcomes and evaluating surrogate endpoints. (2019).

681

682 **Figure legends**

683 Figure 1: PRISMA flow diagram for study inclusion

684 Figure 2: Correlation (r or r_s) between absolute (individual-level) values of ORR and PFS

685 Figure 3: Correlation (r or r_s) between absolute (individual-level) values of ORR and OS

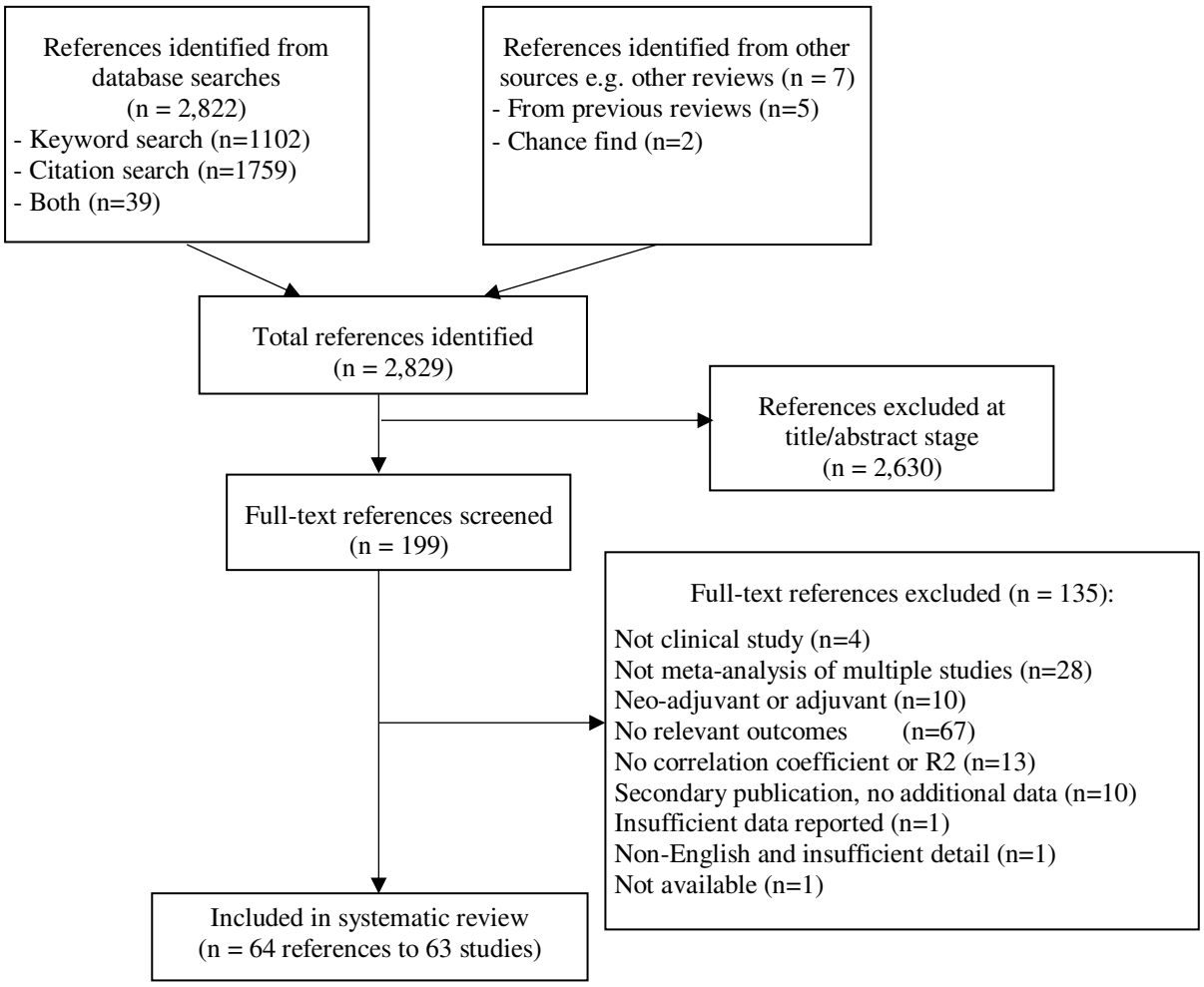
686 Figure 4: Regression R^2 between treatment effects (trial-level) for ORR and PFS

687 Figure 5: Regression R^2 between treatment effects (trial-level) for ORR and OS

688

689

690 **Figure 1: PRISMA flow diagram for study inclusion**



691

692

693 **Table 1: Summary of absolute correlation coefficients and treatment effect R^2 values**

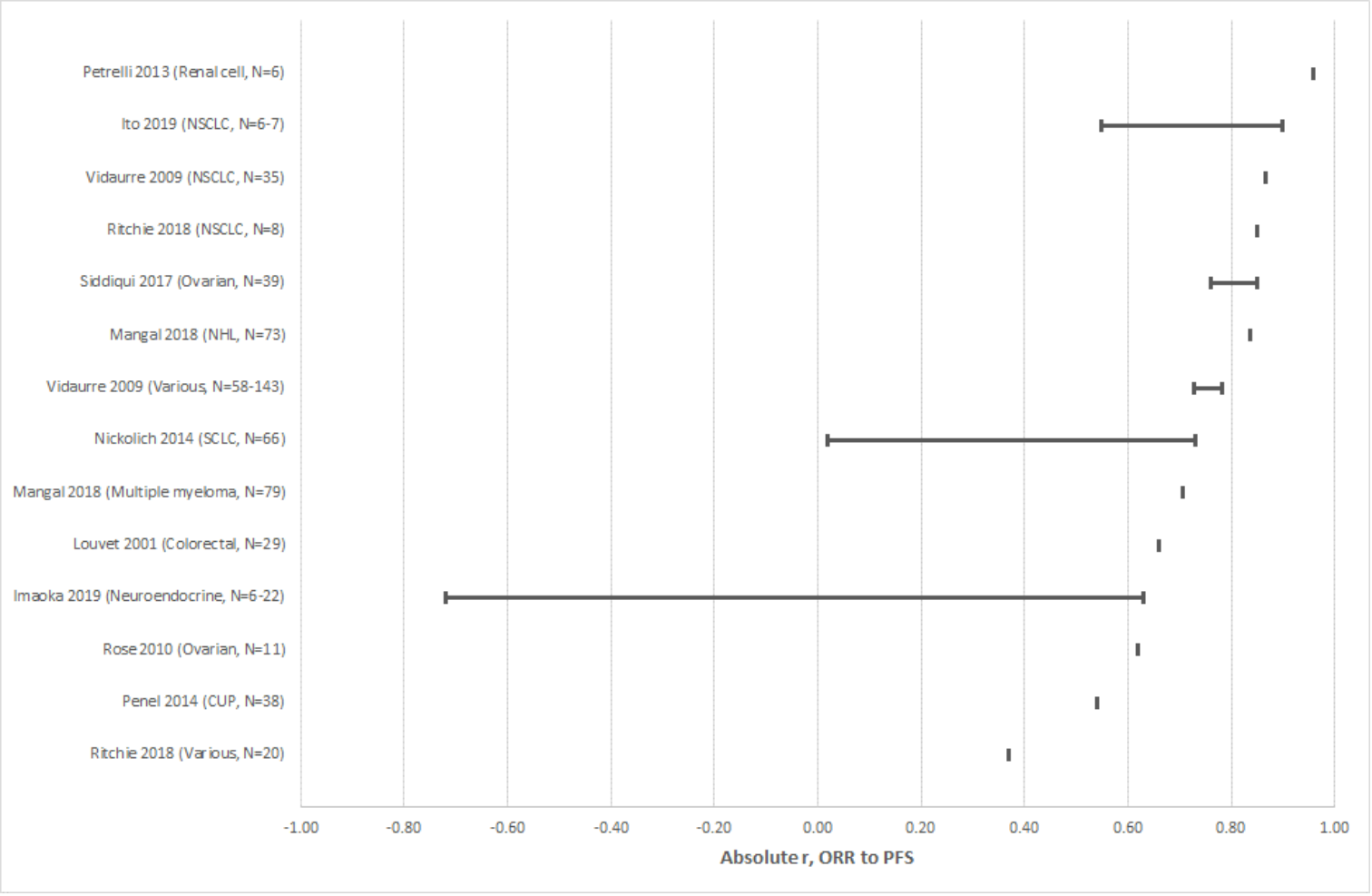
Surrogate relationship	Range of absolute (individual-level) correlations			Range of treatment effect (trial-level) R^2 values		
	N studies	Cancer types and refs	Range of r or r_s across studies and subgroup analyses	N studies	Cancer types and refs	Range of R^2 across studies and subgroup analyses
ORR to PFS	12	NSCLC,[45,65,78] ovarian,[66,72] RCC,[63] NHL,[54] SCLC,[59] MM,[55] CRC,[52] CUP,[62] NET,[44] various[65,78]	-0.72 to 0.96	9	NSCLC,[21,22,45,67,77] ovarian,[27,72] various,[67,77,79] CRC[26,77]	0.18 to 0.94
ORR to TTP	1	Gastric[42]	0.41 to 0.56	0		-
ORR to OS	27	NSCLC,[49,50,65,68,71,45,78] CRC,[35,52,75] ovarian,[66,72] breast,[51,64] gastric,[42,70] various,[65,60,78] pancreatic,[37] RCC,[18,63] gastroesophageal,[61] urothelial,[18,19] AML,[20] SCLC,[59] glioblastoma,[38] CUP,[62] NET[43]	-0.40 to 1.00	31	NSCLC,[21,22,39,40,45,46,58,67,77] CRC,[25,26,29,31,46,73,77] various,[47,57,60,67,77,79] pancreatic,[28,37,53] SCLC,[34,41] RCC,[32,63] breast,[23,36] ovarian,[27] prostate,[30] BTC,[56] STC[74]	-0.08 to 0.84
CR to PFS	2	SCLC,[59] NHL[81]	0.22 to 0.83	1	NHL[69]	0.45 to 0.93
CR to OS	3	NSCLC,[49] SCLC,[59] gastroesophageal[61]	-0.04 to 0.62	2	Breast,[36] SCLC[34]	0.05 to 0.48
PR to PFS	1	SCLC[59]	0.35 to 0.70	0		-
PR to OS	1	SCLC[59]	0.29 to 0.66	0		-
VGPR/CR to PFS	0		(see footnote)*	0		-
DoR to PFS	0		-	0		-
DoR to OS	0		-	0		(see footnote)**

Notes: Further detail on all studies and outcomes is shown in Appendix 5 and Appendix 6. *One study of MM reported the association between VGPR/CR and PFS as adjusted R^2=0.64, but this could not be converted to r because it was adjusted.[55] **Two studies in CRC[29] and pancreatic cancer[28] reported Spearman correlation coefficients between DoR and OS ranging from 0.40 to 0.76, but these could not be converted to R^2 as no Pearson correlation coefficients were reported.

AML, acute myeloid leukaemia; BTC, biliary tract cancer; CR, complete response; CRC, colorectal cancer; CUP, cancer of unknown primary; DoR, duration of response; MM, multiple myeloma; NET, neuroendocrine tumour; NHL, non-Hodgkin's lymphoma; NSCLC, non-small cell lung cancer; ORR, overall response rate; OS, overall survival; PFS, progression-free survival; PR, partial response; RCC, renal cell carcinoma; SCLC, small cell lung cancer; STC, soft tissue sarcoma; TTP, time to progression; VGPR, very good partial response.

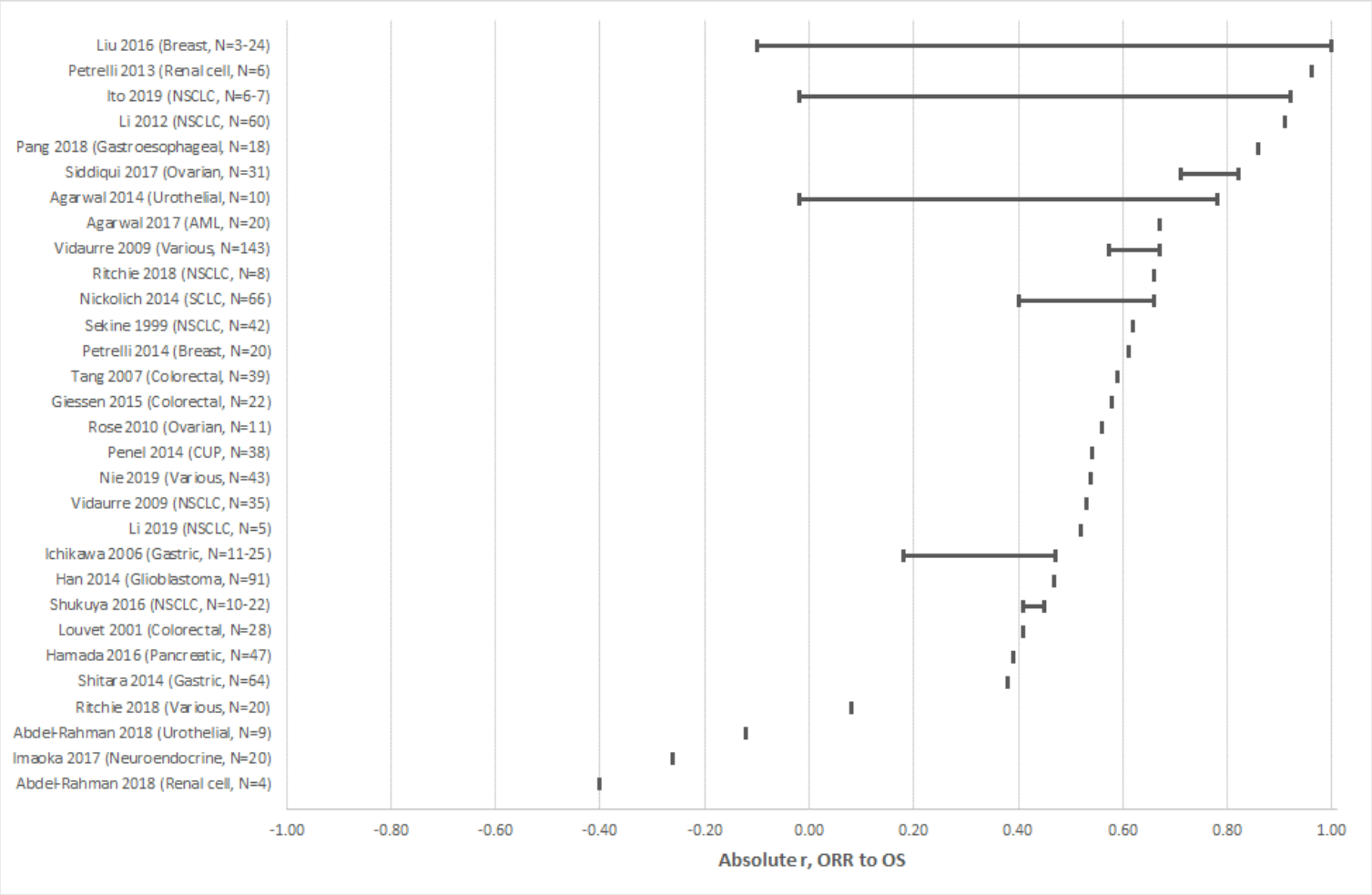
694

Figure 2: Correlation (r or rs) between absolute (individual-level) values of ORR and PFS



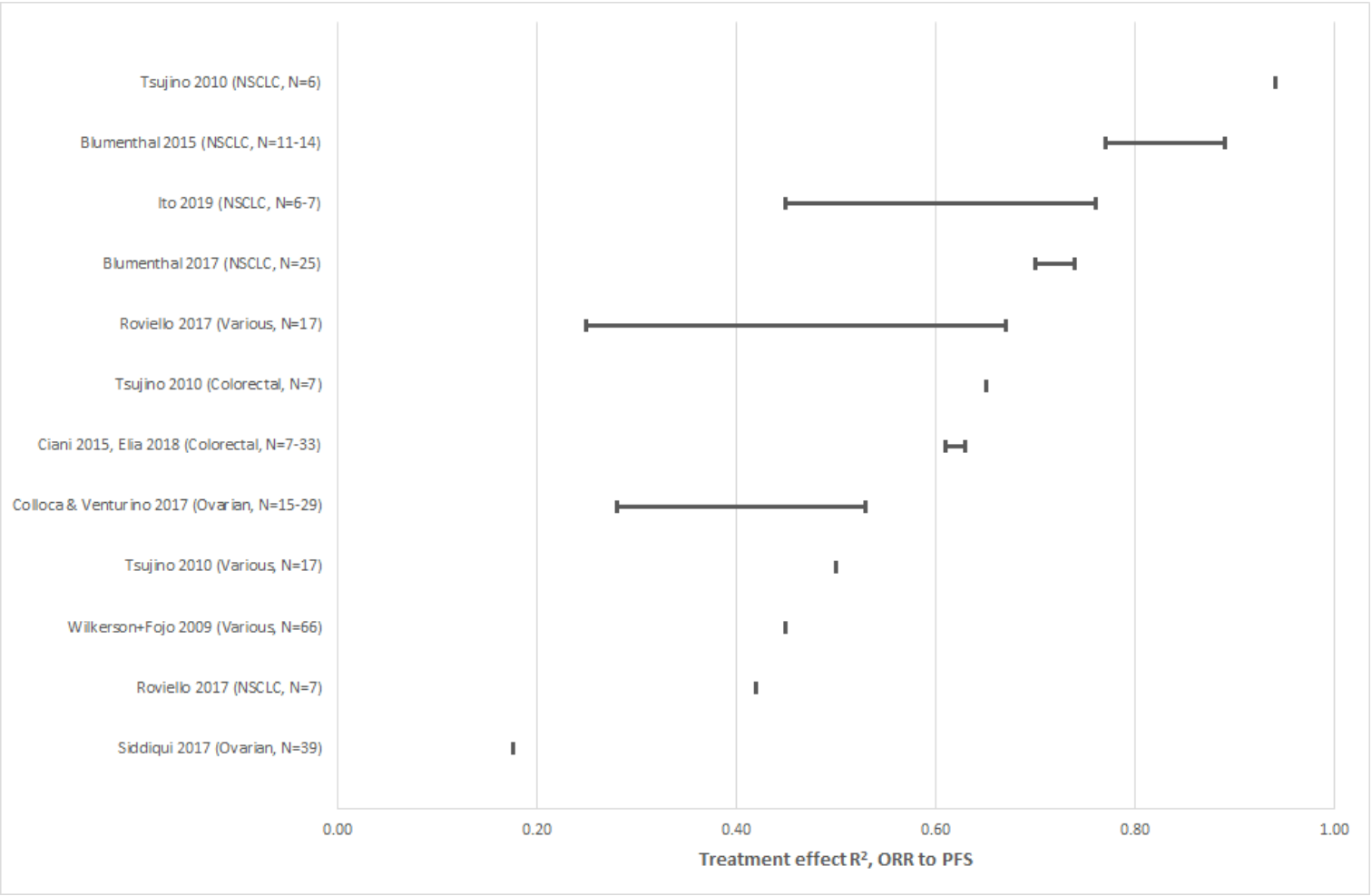
For each study, the plot illustrates the range of correlation coefficients across all subgroup analyses. N represents the number of studies included in each meta-regression. CUP, cancer of unknown primary; NHL, non-Hodgkin's lymphoma; NSCLC, non-small cell lung cancer; ORR, overall response rate; PFS, progression-free survival; SCLC, small cell lung cancer.

Figure 3: Correlation (r or rs) between absolute (individual-level) values of ORR and OS



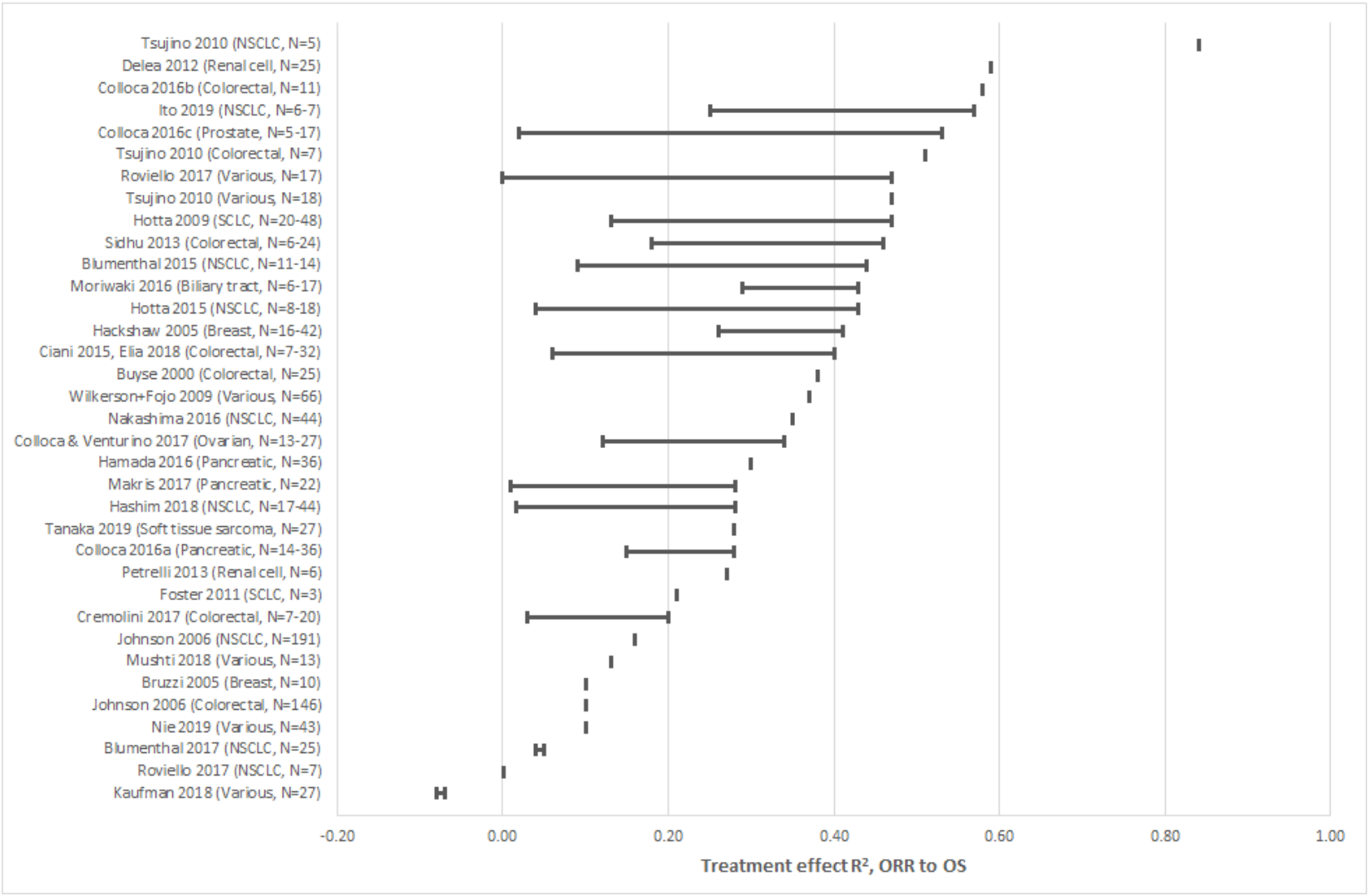
For each study, the plot illustrates the range of correlation coefficients across all subgroup analyses. N represents the number of studies included in each meta-regression. AML, acute myeloid leukaemia; CUP, cancer of unknown primary; NSCLC, non-small cell lung cancer; ORR, overall response rate; OS, overall survival; SCLC, small cell lung cancer.

Figure 4: Regression R2 between treatment effects (trial-level) for ORR and PFS



For each study, the plot illustrates the range of correlation coefficients across all subgroup analyses. N represents the number of studies included in each meta-regression.
NSCLC, non-small cell lung cancer; ORR, overall response rate; PFS, progression-free survival.

Figure 5: Regression R2 between treatment effects (trial-level) for ORR and OS



For each study, the plot illustrates the range of correlation coefficients across all subgroup analyses. N represents the number of studies included in each meta-regression.
NSCLC, non-small cell lung cancer; ORR, overall response rate; OS, overall survival; SCLC, small cell lung cancer.

A systematic review of meta-analyses assessing the validity of tumour response endpoints as surrogates for progression-free or overall survival in cancer

Katy Cooper, Paul Tappenden, Anna Cantrell, Kate Ennis

Supplementary Information **Page**

Supplementary Information 1: MEDLINE search strategy

Search Strategy (March 2019):

1 *Neoplasms/
2 (cancer$ or neoplasm$ or tumour$ or tumour$ or malignan$ or oncology or lymphoma$ or sarcoma$ or melanoma$ or myeloma$ or carcinoma$).tw.
3 1 or 2
4 tumour response$.tw.
5 tumour response$.tw.
6 objective response$.tw.
7 ORR.tw.
8 "duration of response$".tw.
9 dor.tw.
10 response rate$.tw.
11 complete response$.tw
12 overall response$.tw
13 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12
14 3 and 13
15 Regression analysis/
16 regression.tw.
17 relationship.tw.
18 correlation.tw.
19 prediction.tw.
20 association.tw.
21 15 or 16 or 17 or 18 or 19 or 20
22 14 and 21
23 endpoint$.tw.
24 end point$.tw.
25 (surrogate or surrogacy).tw.
26 23 or 24 or 25
27 22 and 26
28 progression-free survival/
29 "progression free survival".tw.
30 "overall survival".tw.
31 (pfs or os).tw.
32 "time to progression".tw.
33 ttp.tw.
34 28 or 29 or 30 or 31 or 32 or 33
35 27 and 34
36 limit 35 to (english language and humans)

Supplementary Table 1: IQWiG scoring criteria

IQWiG Score	Criteria (based on r for treatment-effect association)*
High	Lower confidence interval of r is ≥ 0.85
Medium+	$r \geq 0.85$ with no reported confidence interval **or** $r \geq 0.85$ with wide confidence intervals (lower limit <0.85)
Medium	$0.85 > r \geq 0.7$ and upper confidence interval of r is ≥ 0.7 and lower confidence interval of r is < 0.85, **or** $0.85 > r \geq 0.7$ with no reported confidence interval
Low	Upper confidence interval of r is < 0.7 **or** $r < 0.7$ with no reported confidence interval

Notes:
Based on the scoring criteria reported by IQWiG (2011).[10]
*r is defined as any correlation parameter for the treatment-effect association, e.g. Pearson, Spearman, Kendall's Tau. Where no correlation parameter was reported, if a univariate regression was performed and an R^2 value attained, then r (Pearson correlation coefficient) was calculated as the square-root of R^2. The reported r could be for any treatment effect estimate (hazard ratio, difference in medians, etc.); where more than one was reported, relative estimates (e.g. hazard ratio, odds ratio) were used in preference to difference in medians. The Medium+ category was based on the approach used in Savina *et al.*[14]

Supplementary Table 2: BSES2 scoring criteria

BSES2 score	Criteria (based on R^2 for both treatment effect and individual-level associations)*
Excellent	R^2 (treatment effect) ≥ 0.6 and R^2 (absolute) ≥ 0.6
Good	R^2 (treatment effect) ≥ 0.4 and R^2 (absolute) ≥ 0.4
Fair	R^2 (treatment effect) ≥ 0.2 and R^2 (absolute) ≥ 0.2
Poor	R^2 (treatment effect) < 0.2 and/or R^2 (absolute) < 0.2

Notes:
Based on the scoring criteria reported by Lassere et al. (2012).[11]
*R^2 is the coefficient of determination for a regression analysis. Where R^2 was not reported, it was calculated as the square of the Pearson correlation coefficient (r), if available. The reported R^2 could be for any treatment effect estimate (hazard ratio, difference in medians, etc.); where more than one was reported, relative estimates (e.g. hazard ratio, odds ratio) were used in preference to difference in medians.

Supplementary Table 3: Summary of study characteristics

Surrogate relationship	N	Cancer type	N	Disease stage	N	Line of treatment	N	Treatment type	N
ORR to OS	57	Lung (NSCLC)	16	Advanced/metastatic	43	1st	23	Chemo	21
ORR to PFS	22	Colorectal	10	Unclear	9	All / various	18	Immune checkpoint inhibitors	9
CR to OS	8	Various solid	8	Advanced, locally advanced, unresectable or metastatic	2	NR	8	Targeted	8
CR to PFS	7	Breast	5			1st + 2nd	5	Various	7
DoR to OS	2	NHL	4	Extensive disease	2	2nd	4	Systemic	5
ORR to TTP	1	Lung (SCLC)	3	Limited or extensive disease	1	2nd + subsequent	3	Chemo or targeted	3
PR to PFS	1	Ovarian	3	Advanced or recurrent	1	2nd + 3rd	2	Chemo, immune or targeted	2
PR to OS	1	Pancreatic	3	Advanced, locally advanced or recurrent	1			NR	1
VGPR/CR to PFS	1	Renal cell	3					Chemo + targeted	1
DoR to PFS	1	Gastric	2	Relapsed / refractory	1			Chemo or immune	1
		Neuroendocrine	2	Most stage III/IV	1			Chemo, hormonal + targeted	1
		Soft tissue sarcoma	2	Recurrent / platinum-resistant	1			Chemo or biologic	1
		Urothelial	2	Various	1			Cytokine or targeted	1
		AML	1					Gemcitabine + chemo or targeted	1
		Biliary tract	1						
		Gastroesophageal	1					Bevacizumab + chemo	1
		Glioblastoma	1						
		Multiple myeloma	1						
		Prostate	1						
		Unknown primary	1						

Note: Ns may sum to more than total number of studies (N=63) as some studies reported more than one surrogate relationship or cancer type.
AML, acute myeloid leukaemia; chemo, chemotherapy; CR, complete response; DoR, duration of response; immune, immunotherapy; NR, not reported; NSCLC, non-small cell lung cancer; ORR, overall response rate; OS, overall survival; PFS, progression-free survival; PR, partial response; SCLC, small cell lung cancer; TTP, time to progression; VGPR, very good partial response.

Supplementary Table 4: Summary of reported data types

N primary studies per meta-regression (range)	N patients per meta-regression (range)	Included study types per meta-regression	Data types	Absolute association reported?	Treatment effect association reported?	STE reported?
4 to 191	407 to 44,125	RCT only (N=44) RCT+SA (N=17) SA only (N=2)	AD (N=58) IPD (N=5)	N=32	N=38	N=4

AD, aggregate data; IPD, individual patient data; RCT, randomised controlled trials; SA, single-arm studies; STE, surrogate threshold effect.

5 **Supplementary Table 5: Study characteristics by study**

Reference	Cancer	Surrogate outcome	Final outcome	Stage	Line	Treatment	N studies	N patients	Study types	Publication/ search years	Data type	Response criteria	Absolute association	Treatment effect association	STE reported
Agarwal 2017[20]	Acute myeloid leukemia	ORR CR	OS	Various	1st	Systemic	20†	NR	RCT + SA	2004-2016	AD	NR	Y		
Moriwaki 2016[56]	Biliary tract	ORR	OS	Advanced	1st	Chemo	17†	2040	RCT	Up to 2015	AD	NR		Y	
Bruzzi 2005[23]	Breast	ORR	OS	Metastatic	All	Chemo	10	2126	RCT	1991-2001	IPD	WHO (8), ECOG (1), NR (1)		Y	
Burzykowski 2008[24]	Breast	ORR	PFS OS	Metastatic	1st	Chemo	11	3953	RCT	1999-2008	IPD	WHO		Y	
Hackshaw 2005[36]	Breast	ORR CR	OS	Metastatic	1st	Chemo	42*	9163	RCT	1966-2005	AD	NR		Y	
Liu 2016[51]	Breast	ORR*	OS	Metastatic	2nd + 3rd	Chemo	24	8617	RCT	1999 to 2014	AD	NR	Y		
Petrelli 2014[64]	Breast	ORR	OS	Metastatic or advanced	1st	Targeted + chemo	20†	10138†	RCT	2000 to 2012	AD	NR	Y		
Buyse 2000[25]	Colorectal	ORR	OS	Advanced	1st	Chemo	25	3791	RCT	Collected 1990-1996	IPD	WHO		Y	
Ciani 2015[26] Elia 2018[33]	Colorectal	ORR	PFS OS	Advanced or metastatic	All	Systemic	33	NR	RCT	2003-2013	AD	RECIST or WHO		Y	Y
Colloca 2016b[29]	Colorectal	ORR DoR	OS	Metastatic	1st	Bevacizumab + chemo	11	NR	RCT	2000-2014	AD	RECIST		Y	
Giessen 2015[35]	Colorectal	ORR	OS	Metastatic	2nd	Chemo	22	10509	RCT	2000-2013	AD	RECIST (17), WHO (5)	Y		
Cremolini 2017[31]	Colorectal	ORR	OS	Metastatic	2nd	Targeted	20*	7571	RCT	To 2015	AD	NR		Y	
Johnson 2006[46]	Colorectal	ORR	OS	Metastatic	1st	Chemo	146†	35337†	RCT	To 2005	AD	NR (very few RECIST)		Y	
Louvet 2001[52]	Colorectal	ORR	PFS OS	Metastatic	1st	Various	29	13498	RCT	1990 to 2000	AD	NR	Y		
Sidhu 2013[73]	Colorectal	ORR	OS	Metastatic	1st (most)	Chemo +/- targeted	24†	20438†	RCT	2000 to 2011	AD	NR		Y	
Tang 2007[75]	Colorectal	ORR	OS	Metastatic	1st	Chemo	39	18668	RCT	1990 to 2005	AD	NR	Y	Y	
Tsujino 2010[77]	Colorectal	ORR	PFS OS	Advanced	NR	Targeted	7	NR	RCT	Up to 2009	AD	NR		Y	
Ichikawa 2006[42]	Gastric	ORR	TTP OS	Advanced	1st	Chemo	25	4593	RCT	NR	AD	WHO, SWOG, RECIST, Japan	Y		
Shitara 2014[70]	Gastric	ORR*	OS	Advanced	2nd + 3rd	Chemo	64	4286	RCT + SA	2002 to 2012/2013	AD	NR	Y		

Reference	Cancer	Surrogate outcome	Final outcome	Stage	Line	Treatment	N studies	N patients	Study types	Publication/ search years	Data type	Response criteria	Absolute association	Treatment effect association	STE reported
Pang 2018[61]	Gastroesophageal	ORR* CR	OS	Advanced	1st + 2nd	Targeted	18	7892	RCT	Up to 2018	AD	RECIST	Y		
Han 2014[38]	Glioblastoma	ORR	OS	Unclear	Various	Various	91†	7125†	RCT + SA	1991-2012	AD	NR ("standard criteria")	Y		
Blumenthal 2017[22]	Lung (NSCLC)	ORR	PFS OS	Metastatic	Various	Chemo, immune or targeted	25	20013†	RCT	2003-2016	AD	RECIST or WHO		Y	
Blumenthal 2015[21]	Lung (NSCLC)	ORR	PFS OS	Metastatic	Various	Chemo or targeted	14	12567†	RCT	2003-2013	AD	RECIST (11) or WHO (3)		Y	
Hashim 2018[39]	Lung (NSCLC)	ORR	OS	Advanced	2nd + subsequent	Various	140	41725	RCT	To 2016	AD	NR		Y	Y
Hotta 2015[40]	Lung (NSCLC)	ORR	OS	Advanced	Various	Targeted	18	7633†	RCT	2003-2014	AD	NR		Y	
Ito 2019[45]	Lung (NSCLC)	ORR	PFS OS	Advanced	Various	Immune checkpoint inhibitors (PD-(L)1)	7	3752†	RCT	NR	AD	NR	Y	Y	
Johnson 2006[46]	Lung (NSCLC)	ORR	OS	Advanced	1st	Chemo	191†	44125†	RCT	To 2005	AD	NR (very few RECIST)		Y	
Li 2019[49]	Lung (NSCLC)	ORR* CR	OS	Advanced	1st + 2nd	Immune checkpoint inhibitors	5†	4803†	RCT	Up to 2018	AD	RECIST	Y		
Li 2012[50]	Lung (NSCLC)	ORR	OS	Advanced	1st + 2nd	Targeted	60	9903	RCT + SA	Up to 2011	AD	RECIST (52), WHO (10)	Y		
Nakashima 2016[58]	Lung (NSCLC)	ORR	OS	Advanced, locally advanced and recurrent	1st	Chemo	44	22709	RCT	2005 to 2015	AD	RECIST		Y	
Ritchie 2018[65]	Lung (NSCLC)	ORR*	PFS OS	Advanced	All	Immune checkpoint inhibitors (PD-(L)1 or CTLA4)	8	NR	RCT	2000 to 2017	AD	NR	Y	Y	
Roviello 2017[67]	Lung (NSCLC)	ORR	PFS OS	Unclear	Various	Immune checkpoint inhibitors	7*	3369*	RCT	Up to 2017	AD	RECIST or mWHO		Y	
Sekine 1999[68]	Lung (NSCLC)	ORR	OS	Unclear	Various	Chemo	42	1935	SA +1 RCT	1988-1997	AD	WHO	Y		

Reference	Cancer	Surrogate outcome	Final outcome	Stage	Line	Treatment	N studies	N patients	Study types	Publication/ search years	Data type	Response criteria	Absolute association	Treatment effect association	STE reported
Shukuya 2016[71]	Lung (NSCLC)	ORR	OS	Advanced	All	a) Immune checkpoint inhibitors (PD-(L)1) b) Chemo (docetaxel)	a) 10† b) 22†	NR	RCT + SA	2012 to 2016	AD	RECIST (most)	Y		
Tsujino 2010[77]	Lung (NSCLC)	ORR	PFS OS	Advanced	NR	Targeted	6	NR	RCT	Up to 2009	AD	NR		Y	
Tsujino 2009[76]	Lung (NSCLC)	ORR	PFS OS	Advanced	NR	Targeted	28	6171	RCT + SA	To 2007	AD	RECIST (21), WHO (9)	Y		
Vidaurre 2009[78]	Lung (NSCLC)	ORR*	PFS OS	Advanced, locally advanced, unresectable or metastatic	NR	Chemo or targeted	35	NR	RCT + SA	2006 to 2008	AD	NR	Y		
Foster 2011[34]	Lung (SCLC)	ORR CR	OS	Extensive-stage	1st	Chemo	3 RCTs (32 centres)	596†	RCT	Trials initiated 1987-1999	AD	NR (CR=disappearance; PR ≥50% reduction		Y	
Hotta 2009[41]	Lung (SCLC)	ORR	OS	Extensive disease	1st	Chemo	48	8779	RCT	1990-2008	AD	WHO (23), ECOG (2), RECIST (1), Japan (1), or NR		Y	
Nickolich 2014[59]	Lung (SCLC)	ORR CR PR	PFS OS	Limited or extensive disease	1st + 2nd + maintenance	Various	66†	8471†	RCT + SA	1983 to 2010	AD	NR	Y		
Mangal 2018[55] (myeloma)	Multiple myeloma	ORR* CR VGPR or CR	PFS	Relapsed / refractory	2nd + subsequent	Various	79†	13322†	RCT + SA	1999 to 2016	AD	IMWG	Y		
Imaoka 2019[44]	Neuroendocrine	ORR	PFS	Advanced	Various	Systemic	22	1310	RCT + SA	1996-2016	AD	RECIST (20), WHO (2)	Y		
Imaoka 2017[43]	Neuroendocrine	ORR	OS	Advanced	Various	Systemic	20	2530	RCT + SA	1996-2016	AD	NR	Y		
Lee 2011[48]	NHL (aggressive)	CR	PFS OS	Unclear	1st	Chemo	36†	16103†	RCT	1990-2009	AD	NR		Y	
Lee 2011[48]	NHL (indolent)	CR	PFS OS	Unclear	1st	Chemo	15†	5128†	RCT	1990-2009	AD	NR		Y	

Reference	Cancer	Surrogate outcome	Final outcome	Stage	Line	Treatment	N studies	N patients	Study types	Publication/ search years	Data type	Response criteria	Absolute association	Treatment effect association	STE reported
Mangal 2018[54] (NHL)	NHL	ORR* CR	PFS	Stage III/IV >75% in most cohorts	Various	Various	73	6071	RCT + SA	1996 to 2015	AD	NR	Y		
Shi 2017[69]	NHL (indolent; follicular)	CR 30mo CR 24mo	PFS	Unclear	1st	Chemo or immuno (induction or maintenance)	13	3837	RCT	1990 to 2011	IPD	NR (CR= disappearance)		Y	Y
Zhu 2017[81]	NHL (indolent; follicular)	CR	PFS	Unclear	NR	Chemo, immune or targeted	13	NR	RCT + SA	1993 to 2013	AD	NR	Y		
Zhu 2017[81]	NHL (mantle cell)	CR	PFS	Unclear	NR	Chemo, immune or targeted	NR	NR	RCT + SA	1993 to 2013	AD	NR	Y		
Colloca & Venturino 2017[27]	Ovarian	ORR CR	PFS OS	Advanced	1st	Chemo	29	NR	RCT	1990-2016	AD	WHO (24), RECIST (8)		Y	
Rose 2010[66]	Ovarian	ORR*	PFS OS	Recurrent / platinum-resistant	2nd	Various	11	407	SA	1994 to 2004	IPD	WHO (10), RECIST (1)	Y		
Siddiqui 2017[72]	Ovarian	ORR*	PFS OS	Advanced, recurrent	2nd + subsequent	Chemo	39†	9223†	RCT	2000 to 2015	AD	NR	Y	Y	
Colloca 2016a[28]	Pancreatic	ORR DoR	PFS OS	Advanced or metastatic	1st	Gemcitabine + chemo or targeted	36*	NR	RCT	1997-2014	AD	RECIST		Y	
Hamada 2016[37]	Pancreatic	ORR	OS	Advanced	1st	Chemo	47	15906†	RCT	1995-2015	AD	NR	Y	Y	
Makris 2017[53]	Pancreatic (adenocarcinoma)	ORR	OS	Locally advanced, unresectable or metastatic	1st	Chemo (gemcitabine)	22*	10379*	RCT	2000 to 2015	AD	NR (RR=shrinkage or disappearance)		Y	
Colloca 2016c[30]	Prostate	ORR	OS	Metastatic (castration-resistant)	1st + 2nd	Chemo, hormonal + targeted	17	NR	RCT	1995-2014	AD	NR (CR=disappearance; PR=≥30% reduction)		Y	
Abdel-Rahman 2018[18]	Renal cell	ORR	OS	Advanced	Various	Immune checkpoint inhibitors (PD-(L)1)	4	1093	RCT + SA	To 2017	AD	RECIST	Y		

Reference	Cancer	Surrogate outcome	Final outcome	Stage	Line	Treatment	N studies	N patients	Study types	Publication/ search years	Data type	Response criteria	Absolute association	Treatment effect association	STE reported
Delea 2012[32]	Renal cell	ORR	OS	Metastatic	NR	Cytokine or targeted	25*	10943†	RCT	1997-2010	AD	NR		Y	
Petrelli 2013[63]	Renal cell	ORR	PFS OS	Metastatic	1st	Targeted	6†	3188†	RCT	Up to 2011	AD	NR	Y	Y	
Tanaka 2019[74]	Soft tissue sarcoma	ORR	OS	Advanced	1st	Chemo	27†	6156†	RCT	1974 to 2017	AD	NR		Y	
Zer 2016[80]	Soft tissue sarcoma	ORR	OS	Advanced or metastatic	All	Systemic	52†	9762†	RCT	1974 to 2014	AD	NR		Y	
Penel 2014[62]	Unknown primary	ORR*	PFS OS	Unclear	NR	NR	38†	NR	SA	1997 to 2011	AD	RECIST or WHO	Y		
Abdel-Rahman 2018[18]	Urothelial	ORR	OS	Advanced	Various	Immune checkpoint inhibitors (PD-(L)1)	9	1699	RCT + SA	To 2017	AD	RECIST	Y		
Agarwal 2014[19]	Urothelial	ORR	OS	Advanced (operable or metastatic)	2nd	Chemo or biologic	10	560	RCT + SA	NR	AD	RECIST	Y		
Kaufman 2018[47]	Various solid tumours	ORR	OS	Unclear	Various	Immune checkpoint inhibitors +/- chemo	27†	10300†	RCT	2005-2017	AD	RECIST or mWHO		Y	
Mushti 2018[57]	Various solid tumours	ORR*	OS	Unclear	NR	Immune checkpoint inhibitors (PD-(L)1)	13	6722	RCT	2014 to 2016	AD	RECIST		Y	
Nie 2019[60]	Various solid tumours	ORR*	OS	Advanced or recurrent	Various	Immune checkpoint inhibitors (PD-(L)1)	43†	15088†	RCT + SA	Up to 2018	AD	RECIST	Y	Y	
Ritchie 2018[65]	Various solid tumours	ORR*	PFS OS	Advanced	All	Immune checkpoint inhibitors (PD-(L)1 or CTLA4)	20†	10828†	RCT	2000 to 2017	AD	NR	Y	Y	
Roviello 2017[67]	Various solid tumours	ORR	PFS OS	Unclear	Various	Immune checkpoint inhibitors	17†	8994†	RCT	Up to 2017	AD	RECIST or mWHO		Y	
Tsujino 2010[77]	Various solid tumours	ORR	PFS OS	Advanced	NR	Targeted	18	NR	RCT	Up to 2009	AD	NR		Y	Y

Reference	Cancer	Surrogate outcome	Final outcome	Stage	Line	Treatment	N studies	N patients	Study types	Publication/ search years	Data type	Response criteria	Absolute association	Treatment effect association	STE reported
Vidaurre 2009[78]	Various	ORR*	PFS OS	Advanced, locally advanced, unresectable or metastatic	NR	Chemo or targeted	143†	6974†	RCT + SA	2006 to 2008	AD	NR	Y		
Wilkerson+ Fojo 2009[79]	Various solid tumours	ORR	PFS OS	Metastatic	NR	NR	66†	NR	RCT	NR	AD	NR		Y	

Note: Of the 63 included studies (64 refs), 8 references[18,46,48,65,67,77,78,81] appear on 2-3 rows as they report on 2-3 different cancer types. *Calculated from reported data. †Unclear for individual subgroups.

AD, aggregate data; chemo, chemotherapy; CR, complete response; DoR, duration of response; ECOG, Eastern Cooperative Oncology Group; IMWG, International Myeloma Working Group (criteria); IPD, individual patient data; mo, months; mWHO, modified World Health Organisation (criteria); NHL, non-Hodgkin's lymphoma; NR, not reported; NSCLC, non-small cell lung cancer; ORR, overall response rate (ORR=PR+CR); OS, overall survival; PFS, progression-free survival; PR, partial response; RCT, randomised controlled trials; RECIST, Response Evaluation Criteria In Solid Tumours; SA, single-arm studies; SCLC, small cell lung cancer; STE, surrogate threshold effect; TTP, time to progression; VGPR, very good partial response; WHO, World Health Organisation (criteria).

7 **Supplementary Table 6: Quality assessment of included meta-reviews**

Reference	Cancer	Inclusion criteria clear & relevant (population, outcomes, study type)	Literature search comprehensive (at least 2 databases or other sources AND keywords provided)	Duplicate study selection (all or a sample)	Duplicate data extraction or data checking	Risk of bias assessment reported	Analysis methods appropriate: correlation coefficient (r or rs) and/or coefficient of determination (R2)	Heterogeneity explored through subgroup analyses	Uncertainty assessed (reports 95% confidence intervals for r, rs or R2)
Agarwal 2017[20]	AML	Y	N (PubMed only)	U	U	N	Y	N	N
Moriwaki 2016[56]	Biliary tract	Y	Y	U	Y	N	Y	Y	Y
Bruzzi 2005[23]	Breast	Y	Y	U	Y	N	Y	N	Y
Burzykowski 2008[24]	Breast	Y	N	U	U	N	Y	N	Y
Hackshaw 2005[36]	Breast	Y	N (Medline only)	U	U	N	Y	Y	N
Liu 2016[51]	Breast	Y	Y	Y	Y	N	Y	Y	Y
Petrelli 2014[64]	Breast	Y	Y	U	U	N	Y	N	Y
Buyse 2000[25]	Colorectal	Y	U	U	Y	N	Y	N	Y
Ciani 2015[26] Elia 2018[33]	Colorectal	Y	Y	Y	Y	Y	Y	Y	Y
Colloca 2016b[29]	Colorectal	Y	Y	U	U	N	Y	N	N
Giessen 2015[35]	Colorectal	Y	Y	U	U	N	Y	N	Y
Cremolini 2017[31]	Colorectal	Y	Y	U	Y	N	Y	Y	N
Johnson 2006[46]	Colorectal Lung (NSCLC)	Y	Y	N	Y	Y	Y	N	N
Louvet 2001[52]	Colorectal	Y	U	U	U	N	Y	N	N
Sidhu 2013[73]	Colorectal	Y	Y	U	U	N	Y	Y	Y
Tang 2007[75]	Colorectal	Y	Y	U	U	N	Y	N	Y
Tsujino 2010[77]	Colorectal Lung (NSCLC) Various tumours	Y	N (PubMed only)	U	Y	N	Y	N	N
Ichikawa 2006[42]	Gastric	Y	Y	U	U	N	Y	Y	N
Shitara 2014[70]	Gastric	Y	Y	U	U	N	Y	N	Y
Pang 2018[61]	Gastroesophageal	Y	N (search terms NR)	U	Y	Y	Y	N	N
Han 2014[38]	Glioblastoma	Y	Y	U	U	N	Y	N	Y
Blumenthal 2017[22]	Lung (NSCLC)	Y	N (trials submitted to FDA rather than search)	U	U	N	Y	N	Y
Blumenthal 2015[21]	Lung (NSCLC)	Y	N (FDA trials not search)	U	U	N	Y	Y	Y

Reference	Cancer	Inclusion criteria clear & relevant (population, outcomes, study type)	Literature search comprehensive (at least 2 databases or other sources AND keywords provided)	Duplicate study selection (all or a sample)	Duplicate data extraction or data checking	Risk of bias assessment reported	Analysis methods appropriate: correlation coefficient (r or rs) and/or coefficient of determination (R2)	Heterogeneity explored through subgroup analyses	Uncertainty assessed (reports 95% confidence intervals for r, rs or R2)
Hashim 2018[39]	Lung (NSCLC)	Y	Y	Y	Y	Y	Y	Y	Y
Hotta 2015[40]	Lung (NSCLC)	Y	Y	U	Y	N	Y	Y	N
Ito 2019[45]	Lung (NSCLC)	Y	Y	Y	U	N	Y	Y	N
Li 2019[49]	Lung (NSCLC)	Y	N (search terms NR)	U	Y	Y	Y	N	N
Li 2012[50]	Lung (NSCLC)	Y	Y	Y	Y	N	Y	N	N
Nakashima 2016[58]	Lung (NSCLC)	Y	Y	Y	Y	Y	Y	N	N
Ritchie 2018[65]	Lung (NSCLC) Various tumours	Y	Y	U	Y	Y	Y	N	Y
Roviello 2017[67]	Lung (NSCLC) Various tumours	Y	Y	Y	Y	N	Y	Y (for various)	Y
Sekine 1999[68]	Lung (NSCLC)	Y	Y	U	U	N	Y	N	N
Shukuya 2016[71]	Lung (NSCLC)	Y	Y	U	Y	N	Y	Y	N
Tsujino 2009[76]	Lung (NSCLC)	Y	Y	U	U	N	N (slope only)	N	N
Vidaurre 2009[78]	Lung (NSCLC) Various	Y	N (trials in 5 journals rather than search)	U	Y	N	Y	Y (for various)	N
Foster 2011[34]	Lung (SCLC)	Y	N (trials by 1 group rather than search)	U	U	N	Y	N	N
Hotta 2009[41]	Lung (SCLC)	Y	Y	U	Y	N	Y	Y	N
Nickolich 2014[59]	Lung (SCLC)	Y	N (trials in 1 journal rather than search)	U	Y	N	Y	Y	N
Mangal 2018[55] (myeloma)	Multiple myeloma	Y	Y	U	U	N	Y	N	N
Imaoka 2019[44]	Neuroendocrine	Y	Y	Y	U	N	Y	Y	Y
Imaoka 2017[43]	Neuroendocrine	Y	Y	Y	U	N	Y	N	Y
Lee 2011[48]	NHL (aggressive) NHL (indolent)	Y	Y	U	U	N	Y	N	Y
Mangal 2018[54] (NHL)	NHL	Y	Y	U	U	N	Y	N	N
Shi 2017[69]	NHL (follicular)	Y	Y	U	U	N	Y	Y	Y
Zhu 2017[81]	NHL (follicular) NHL (mantle cell)	Y	Y	U	U	N	Y	N	Y

Reference	Cancer	Inclusion criteria clear & relevant (population, outcomes, study type)	Literature search comprehensive (at least 2 databases or other sources AND keywords provided)	Duplicate study selection (all or a sample)	Duplicate data extraction or data checking	Risk of bias assessment reported	Analysis methods appropriate: correlation coefficient (r or rs) and/or coefficient of determination (R2)	Heterogeneity explored through subgroup analyses	Uncertainty assessed (reports 95% confidence intervals for r, rs or R2)
Colloca & Venturino 2017[27]	Ovarian	Y	Y	Y	U	N	Y	Y	N
Rose 2010[66]	Ovarian	Y	N (trials by 1 group rather than search)	N	U	N	Y	N	N
Siddiqui 2017[72]	Ovarian	Y	Y	U	Y	N	Y	N	N
Colloca 2016a[28]	Pancreatic	Y	N (PubMed only)	Y	U	N	Y	Y	N
Hamada 2016[37]	Pancreatic	Y	Y	Y	Y	N	Y	N	Y
Makris 2017[53]	Pancreatic (adenocarcinoma)	Y	N (search terms NR)	Y	Y	N	Y	N	Y
Colloca 2016c[30]	Prostate	Y	N (PubMed only)	Y	Y	N	Y	Y	N
Abdel-Rahman 2018[18]	Renal cell Urothelial	Y	Y	U	U	N	Y	N	N
Delea 2012[32]	Renal cell	Y	N (search terms NR)	Y	U	N	Y	N	N
Petrelli 2013[63]	Renal cell	Y	Y	U	Y	N	Y	N	N
Tanaka 2019[74]	Soft tissue sarcoma	Y	Y	U	Y	N	Y	N	Y
Zer 2016[80]	Soft tissue sarcoma	Y	Y	U	Y	Y	Y	N	N
Penel 2014[62]	Unknown primary	Y	N (Medline only; search terms NR)	U	U	N	Y	N	N
Agarwal 2014[19]	Urothelial	Y	N (search methods NR)	U	U	N	Y	Y	N
Kaufman 2018[47]	Various tumours	Y	Y	Y	Y	N	Y	Y	N
Mushti 2018[57]	Various tumours	Y	N (FDA trials not search)	U	U	N	Y	N	N
Nie 2019[60]	Various tumours	Y	Y	Y	Y	N	Y	N	N
Wilkerson+Fojo 2009[79]	Various tumours	Y	N (search methods NR)	U	Y	N	Y	N	N

AML, acute myeloid leukaemia; N, No; NHL, non-Hodgkin's lymphoma; NR, not reported; NSCLC, non-small cell lung cancer; SCLC, small cell lung cancer; U, Unclear; Y; Yes.

Supplementary Table 7: Absolute correlation and regression results per study

Ref	SO	FO	Cancer	Line Sub-groups	Treatment	N stds	N pts	Absolute correlation methods	Correlation coefficient (95% CI), p-value	Absolute regression Methods	Regression R2 (95% CI), p-value	Linear regression equation
ORR vs. PFS (or TTP)												
Louvet 2001[52]	ORR	PFS	Colorectal	1st	Various	29	13498	Spearman (ORR vs. med PFS)	rs=0.66, p<0.0001	LR (ORR vs. med PFS)		PFS = 3.2 + 0.1 * ORR
Ichikawa 2006[42]	ORR	TTP	Gastric	1st	Chemo (any)	12*	2144	Spearman, wtd O(RR vs. med TTP)	rs=0.49, p<0.0001	WLR (ORR vs. med TTP)		TTP = 1.73 + 0.09 * ORR
Ichikawa 2006[42]	ORR	TTP	Gastric	1st	Chemo (novel)	8*	1077	Spearman, wtd (ORR vs. med TTP)	rs=0.41, p=0.018			
Ichikawa 2006[42]	ORR	TTP	Gastric	1st	Chemo (non-novel)	7*	1067	Spearman, wtd (ORR vs. med TTP)	rs=0.56, p=0.0053			
Ito 2019[45]	ORR	PFS	Lung (NSCLC)	Various	Immune checkpoint inhibitors (PD-(L)1)	6	3752†	a) Pearson, wtd b) Spearman, wtd (ORR vs. med PFS)	a) r=0.55, p<0.0001 b) rs=0.33, p<0.0001	WLR R2 (ORR vs. med PFS)	R2=0.30, p=0.206	
Ito 2019[45]	ORR	PFS	Lung (NSCLC)	- Various - High PD-L1 expression	Immune checkpoint inhibitors (PD-(L)1)	7	1381	a) Pearson, wtd b) Spearman, wtd (ORR vs. med PFS)	a) r=0.90, p<0.0001 b) rs=0.48, p<0.0001	WLR R2 (ORR vs. med PFS)	R2=0.81, p=0.006	
Ritchie 2018[65]	ORR	PFS	Lung (NSCLC)	All	Immune checkpoint inhibitors (PD-(L)1 or CTLA4)	8	NR	Correlation (NR) (ORR vs. 6mo PFS)	r=0.85 (0.63 to 1.06), p=NR			
Tsujino 2009[76]	ORR	PFS	Lung (NSCLC)	NR	Targeted	18*	3790*			LR (ORR vs. med PFS)	R2=NR, p=0.001	Slope 0.072
Vidaurre 2009[78]	ORR	PFS	Lung (NSCLC)	NR	Chemo or targeted	35	NR			Regression (NR) (ORR vs. med PFS)	R2=0.75, p<0.0001	
Nickolich 2014[59]	ORR	PFS	Lung (SCLC)	- 1st + 2nd + maintenance - Limited or extensive	Various	66†	8471†	Pearson (ORR vs. med PFS)	r=0.73, p<0.0001			
Nickolich 2014[59]	ORR	PFS	Lung (SCLC)	- 1st + 2nd + maintenance - Limited disease	Various	66†	8471†	Pearson (ORR vs. med PFS)	r=0.02, p=0.978			
Nickolich 2014[59]	ORR	PFS	Lung (SCLC)	- 1st + 2nd + maintenance - Extensive disease	Various	66†	8471†	Pearson (ORR vs. med PFS)	r=0.51, p=0.013			

Ref	SO	FO	Cancer	Line Sub-groups	Treatment	N stds	N pts	Absolute correlation methods	Correlation coefficient (95% CI), p-value	Absolute regression Methods	Regression R2 (95% CI), p-value	Linear regression equation
Mangal 2018[55] (myeloma)	ORR	PFS	Multiple myeloma	2nd +	Various	79†	13322†			WLR adj R2 (logit ORR vs. log med PFS)	Adj R2=0.50, p=NR	
Imaoka 2019[44]	ORR	PFS	Neuroendocrine	Various	Systemic	22	1310	Pearson (ORR vs. med PFS)	r=0.37 (-0.05 to 0.80), p=0.085			
Imaoka 2019[44]	ORR	PFS	Neuroendocrine	- Various - Published 1996-2010	Systemic	6*	NR	Pearson (ORR vs. med PFS)	r= -0.08 (-0.76 to 0.60), p=0.824			
Imaoka 2019[44]	ORR	PFS	Neuroendocrine	- Various - Published 2011-2016	Systemic	16*	NR	Pearson (ORR vs. med PFS)	r=0.43 (-0.07 to 0.93), p=0.095			
Imaoka 2019[44]	ORR	PFS	Neuroendocrine	Various	Cytotoxic	9 arms	NR	Pearson (ORR vs. med PFS)	r=0.63 (0.03 to 1.22), p=0.041			
Imaoka 2019[44]	ORR	PFS	Neuroendocrine	Various	Non-cytotoxic	18 arms	NR	Pearson (ORR vs. med PFS)	r=0.18 (-0.27 to 0.62), p=0.432			
Imaoka 2019[44]	ORR	PFS	Neuroendocrine	Various	Targeted	19 arms	NR	Pearson (ORR vs. med PFS)	r=0.42 (-0.06 to 0.90), p=0.086			
Imaoka 2019[44]	ORR	PFS	Neuroendocrine	Various	Non-targeted	8 arms	NR	Pearson (ORR vs. med PFS)	r= -0.72 (-1.09 to -0.35), p<0.001			
Mangal 2018[54] (NHL)	ORR	PFS	NHL	Various	Various	73	6071			LR adj R2 (logit ORR vs. log med PFS)	Adj R2=0.70, p=NR	log (med PFS) = 1.97 + 0.414 * logit (ORR)
Rose 2010[66]	ORR	PFS	Ovarian	2nd	Various	11	407	a) Pearson b) Kendall Tau-b (ORR vs. med PFS)	a) r=0.62, p=0.044 b) r=0.48, p=0.042			
Siddiqui 2017[72]	ORR	PFS	Ovarian	2nd +	Chemo	39†	9223†	a) Pearson, wtd (ORR vs. med PFS) b) Pearson, unwtd (ORR vs. med PFS)	a) r=0.85, p<0.001 b) 0.76, p<0.001	WLR R2 (ORR vs. med PFS): a) unadj b) adj	a) R2=0.72, p=NR b) adj R2=0.72, p=NR	med PFS = 2.59 + 0.12 * ORR
Petrelli 2013[63]	ORR	PFS	Renal cell	1st	Targeted	6†	3188†	Spearman, wtd (ORR vs. med PFS)	rs=0.96, p<0.0001			
Penel 2014[62]	ORR	PFS	Unknown primary	NR	NR	38†	NR	Pearson via WLR (ORR v. med PFS)	r=0.54, p<0.0001			
Ritchie 2018[65]	ORR	PFS	Various solid tumours	All	Immune checkpoint inhibitors (PD-(L)1 or CTLA4)	20†	10828†	Correlation (NR) (ORR vs. 6mo PFS)	r=0.37 (0.06 to 0.95), p=NR			

Ref	SO	FO	Cancer	Line Sub-groups	Treatment	N stds	N pts	Absolute correlation methods	Correlation coefficient (95% CI), p-value	Absolute regression Methods	Regression R2 (95% CI), p-value	Linear regression equation
Vidaurre 2009[78]	ORR	PFS	Various	NR	Chemo	85	3982*			Regression (NR) (ORR vs. med PFS)	R2=0.53, p<0.0001	
Vidaurre 2009[78]	ORR	PFS	Various	NR	Targeted	58	2992*			Regression (NR) (ORR vs. med PFS)	R2=0.61, p<0.0001	
Vidaurre 2009[78]	ORR	PFS	Various	NR	Chemo or targeted	143†	6974†			Regression (NR) (ORR vs. med PFS)	R2=0.56, p<0.0001	
ORR vs. OS												
Agarwal 2017[20]	ORR	OS	Acute myeloid leukemia	1st	Systemic	20†	NR			WLR adj R2 (logit ORR vs. log med OS)	Adj R2=0.45, p=NR	
Liu 2016[51]	ORR	OS	Breast	2nd + 3rd	Chemo	24	8617	Spearman (ORR vs. med OS)	rs=0.54 (0.29 to 0.72), p<0.0001			
Liu 2016[51]	ORR	OS	Breast	- 2nd + 3rd - Previous anthracycline/taxanes	Chemo	15*	NR	Spearman (ORR vs. med OS)	rs=0.62 (0.32 to 0.84), p=NR			
Liu 2016[51]	ORR	OS	Breast	- 2nd + 3rd - Previous trastuzumab/bevacizumab	Chemo	5*	NR	Spearman (ORR vs. med OS)	rs=0.78 (0.19 to 1.0), p=NR			
Liu 2016[51]	ORR	OS	Breast	2nd + 3rd	Chemo (taxanes)	21*	NR	Spearman (ORR vs. med OS)	rs=0.49 (-0.19 to 0.92), p=NR			
Liu 2016[51]	ORR	OS	Breast	2nd + 3rd	Chemo (antimetabolites)	22*	NR	Spearman (ORR vs. med OS)	rs=-0.10, p=NR			
Liu 2016[51]	ORR	OS	Breast	- 2nd + 3rd - HER2-pos	Chemo	5*	NR	Spearman (ORR vs. med OS)	rs=0.96 (0.80 to 1.00), p=NR			
Liu 2016[51]	ORR	OS	Breast	- 2nd + 3rd - HER2-neg	Chemo	3*	NR	Spearman (ORR vs. med OS)	rs=1.00, p=NR			
Petrelli 2014[64]	ORR	OS	Breast	1st	Targeted + chemo	20†	10138†	Spearman, wtd (ORR vs. med OS)	rs=0.61 (0.59 to 0.63), p=NR			
Giessen 2015[35]	ORR	OS	Colorectal	2nd	Chemo	22	10509	Pearson, wtd (log odds ORR vs. log med OS)	r=0.58 (0.38 to 0.72), p=0.003			
Louvet 2001[52]	ORR	OS	Colorectal	1st	Various	28*	13284*	Spearman (ORR vs. med OS)	rs=0.41, p=0.0009	LR (ORR vs. med OS)		OS = 10.45 + 0.088 * ORR

Ref	SO	FO	Cancer	Line Sub-groups	Treatment	N stds	N pts	Absolute correlation methods	Correlation coefficient (95% CI), p-value	Absolute regression Methods	Regression R2 (95% CI), p-value	Linear regression equation
Tang 2007[75]	ORR	OS	Colorectal	1st	Chemo	39	18668	Spearman (ORR vs. med OS)	rs=0.59 (0.42 to 0.72), p<0.000001			
Ichikawa 2006[42]	ORR	OS	Gastric	1st	Chemo (any)	25	4593	Spearman, wtd (ORR vs. med OS)	rs=0.45, p<0.0001	WLR (ORR vs. med OS)		OS = 5.89 + 0.08 * ORR
Ichikawa 2006[42]	ORR	OS	Gastric	1st	Chemo (novel)	11*	1170	Spearman, wtd (ORR vs. med OS)	rs=0.18, p=0.12			
Ichikawa 2006[42]	ORR	OS	Gastric	1st	Chemo (non-novel)	20*	3423	Spearman, wtd (ORR vs. med OS)	rs=0.47, p<0.0001			
Shitara 2014[70]	ORR	OS	Gastric	2nd + 3rd	Chemo	64	4286	Spearman (ORR vs. med OS)	rs=0.38 (0.16 to 0.6), p=NR			
Pang 2018[61]	ORR	OS	Gastroesophageal	1st + 2nd	Targeted	18	7892	Correlation (NR) (ORR vs. med OS)	r=0.86, p<0.0001			
Han 2014[38]	ORR	OS	Glioblastoma	Various	Various	91†	7125†			WLR R2 (ORR vs. med OS)	R2=0.22 (0.04 to 0.42), p=NR	
Ito 2019[45]	ORR	OS	Lung (NSCLC)	Various	Immune checkpoint inhibitors (PD-(L)1)	6	3752†	a) Pearson, wtd b) Spearman, wtd (ORR vs. med OS)	a) r= -0.02, p=0.4564 b) rs= -0.14, p<0.0001			
Ito 2019[45]	ORR	OS	Lung (NSCLC)	- Various - High PD-L1 expression	Immune checkpoint inhibitors (PD-(L)1)	7	1381	a) Pearson, wtd b) Spearman, wtd (ORR vs. med OS)	a) r=0.92, p<0.0001 b) rs=0.77, p<0.0001	WLR R2 (ORR vs. med OS)	R2=0.84, p=0.004	
Li 2019[49]	ORR	OS	Lung (NSCLC)	1st + 2nd	Immune checkpoint inhibitors	5†	4803†	Pearson (ORR vs. med OS)	r=0.52, p=0.28	LR (ORR vs. med OS)	R2=0.27, p=NR	
Li 2012[50]	ORR	OS	Lung (NSCLC)	1st + 2nd	Targeted	60	9903			WLSR R2 (ORR vs. med OS)	R2=0.83, p<0.000001	
Ritchie 2018[65]	ORR	OS	Lung (NSCLC)	All	Immune checkpoint inhibitors (PD-(L)1 or CTLA4)	8	NR	Correlation (NR) (ORR vs. 12mo OS)	r=0.66 (0.17 to 1.08), p=NR			
Sekine 1999[68]	ORR	OS	Lung (NSCLC)	Various	Chemo	42	1935	Pearson (ORR vs. med OS)	r=0.62, p<0.001			
Shukuya 2016[71]	ORR	OS	Lung (NSCLC)	All	Immune checkpoint inhibitors (PD-(L)1)	10†	NR	Spearman, wtd (ORR vs. med OS)	rs=0.45, p=0.141			
Shukuya 2016[71]	ORR	OS	Lung (NSCLC)	All	Chemo (docetaxel)	22†	NR	Spearman, wtd (ORR vs. med OS)	rs=0.41, p=0.053			

Ref	SO	FO	Cancer	Line Sub-groups	Treatment	N stds	N pts	Absolute correlation methods	Correlation coefficient (95% CI), p-value	Absolute regression Methods	Regression R2 (95% CI), p-value	Linear regression equation
Tsujino 2009[76]	ORR	OS	Lung (NSCLC)	NR	Targeted	28	6171			LR (ORR vs. med OS)	R2=NR, p<0.0001	Slope 0.258
Vidaurre 2009[78]	ORR	OS	Lung (NSCLC)	NR	Chemo or targeted	35	NR			Regression (NR) (ORR vs. med OS)	R2=0.28, p=0.0024	
Nickolich 2014[59]	ORR	OS	Lung (SCLC)	- 1st + 2nd + maintenance - Limited or extensive	Various	66†	8471†	Pearson (ORR vs. med OS)	r=0.66, p<0.0001			
Nickolich 2014[59]	ORR	OS	Lung (SCLC)	- 1st + 2nd + maintenance - Limited disease	Various	66†	8471†	Pearson (ORR vs. med OS)	r=0.40, p=0.193			
Nickolich 2014[59]	ORR	OS	Lung (SCLC)	- 1st + 2nd + maintenance - Extensive disease	Various	66†	8471†	Pearson (ORR vs. med OS)	r=0.44, p=0.012			
Imaoka 2017[43]	ORR	OS	Neuroendocrine	Various	Systemic	20	2530	Spearman (ORR vs. med OS)	rs= -0.26 (-0.64 to 0.11), p=0.164			
Rose 2010[66]	ORR	OS	Ovarian	2nd	Various	11	407	a) Pearson b) Kendall Tau-b (ORR vs. med OS)	a) r=0.56, p=0.071 b) r=0.40, p=0.086			
Siddiqui 2017[72]	ORR	OS	Ovarian	2nd +	Chemo	31†	9223†	a) Pearson, wtd (ORR vs. med OS) b) Pearson, unwtd (ORR vs. med OS)	a) r=0.82, p<0.001 b) 0.71, p<0.001	WLR R2 (ORR vs. med OS): a) unadj b) adj	a) R2=0.67, p=NR b) adj R2=0.66, p=NR	med OS = 9.48 + 0.28 * ORR
Hamada 2016[37]	ORR	OS	Pancreatic	1st	Chemo	47	15906†	Spearman (ORR vs. med OS)	rs=0.39 (0.20 to 0.55), p<0.001			
Abdel-Rahman 2018[18]	ORR	OS	Renal cell	Various	Immune checkpoint inhibitors (PD-(L)1)	4	1093	Pearson (ORR vs. med OS)	r= -0.40, p=0.436			
Petrelli 2013[63]	ORR	OS	Renal cell	1st	Targeted	6†	3188†	Spearman, wtd (ORR vs. med OS)	rs=0.96, p<0.0001			
Penel 2014[62]	ORR	OS	Unknown primary	NR	NR	38†	NR	Pearson via WLR (ORR v. med OS)	r=0.54, p<0.0001			

Ref	SO	FO	Cancer	Line Sub-groups	Treatment	N stds	N pts	Absolute correlation methods	Correlation coefficient (95% CI), p-value	Absolute regression Methods	Regression R2 (95% CI), p-value	Linear regression equation
Abdel-Rahman 2018[18]	ORR	OS	Urothelial	Various	Immune checkpoint inhibitors (PD-(L)1)	9	1699	Pearson (ORR vs. med OS)	r= -0.12, p=0.758			
Agarwal 2014[19]	ORR	OS	Urothelial	2nd	Chemo or biologic	10	560	Pearson (ORR vs. 12mo OS)	r=0.37, p=0.30	WLR R2 (ORR vs. 12mo OS): a) unadj b) adj (RE)	a) R2=0.26, p=NR b) Adj R2=0.16, p=0.1359	
Agarwal 2014[19]	ORR	OS	Urothelial	- 2nd - Operable	Chemo	NR	214†	Pearson (ORR vs. 12mo OS)	r=0.78, p=NR	WLR adj R2 (ORR vs. 12mo OS)	Adj R2=0.54, p=NR	
Agarwal 2014[19]	ORR	OS	Urothelial	- 2nd - Metastatic	Chemo	NR	391†	Pearson (ORR vs. 12mo OS)	r= -0.018, p=NR	WLR adj R2 (ORR vs. 12mo OS)	Adj R2= -0.13, p=NR	
Nie 2019[60]	ORR	OS	Various solid tumours	Various	Immune checkpoint inhibitors (PD-(L)1)	43†	15088†			Squared Spearman (ORR vs. med OS)	r2s=0.29, p<0.001	
Ritchie 2018[65]	ORR	OS	Various solid tumours	All	Immune checkpoint inhibitors (PD-(L)1 or CTLA4)	20†	10828†	Correlation (NR) (ORR vs. 12mo OS)	r=0.08 (-0.17 to 0.70), p=NR			
Vidaurre 2009[78]	ORR	OS	Various	NR	Chemo	85	3982*			Regression (NR) (ORR vs. med OS)	R2=0.35, p<0.0001	
Vidaurre 2009[78]	ORR	OS	Various	NR	Targeted	58	2992*			Regression (NR) (ORR vs. med OS)	R2=0.45, p<0.0001	
Vidaurre 2009[78]	ORR	OS	Various	NR	Chemo or targeted	143†	6794†			Regression (NR) (ORR vs. med OS)	R2=0.33, p<0.0001	
CR vs. PFS												
Nickolich 2014[59]	CR	PFS	Lung (SCLC)	- 1st + 2nd + maintenance - Limited or extensive	Various	66†	8471†	Pearson (CR vs. med PFS)	r=0.71, p<0.0001			
Nickolich 2014[59]	CR	PFS	Lung (SCLC)	- 1st + 2nd + maintenance - Limited disease	Various	66†	8471†	Pearson (CR vs. med PFS)	r=0.22, p=0.491			

Ref	SO	FO	Cancer	Line Sub-groups	Treatment	N stds	N pts	Absolute correlation methods	Correlation coefficient (95% CI), p-value	Absolute regression Methods	Regression R2 (95% CI), p-value	Linear regression equation
Nickolich 2014[59]	CR	PFS	Lung (SCLC)	- 1st + 2nd + maintenance - Extensive disease	Various	66†	8471†	Pearson (CR vs. med PFS)	r=0.35, p=0.116			
Mangal 2018[55] (myeloma)	CR	PFS	Multiple myeloma	2nd +	Various	79†	13322†			WLR adj R2 (logit CR vs. log med PFS)	Adj R2=0.47, p=NR	
Mangal 2018[54] (NHL)	CR	PFS	NHL	Various	Various	73	6071			LR adj R2 (logit CR vs. log med PFS)	Adj R2=0.57, p=NR	log (med PFS) = 2.38 + 0.340 * logit (CR)
Zhu 2017[81]	CR	PFS	NHL (indolent; follicular)	NR	Chemo, immune or targeted	13	NR			WLR R2: a) CR vs. med PFS b) CR vs. 3-year PFS	a) R2=0.69 (0.22 to 0.89), p=NR b) R2=0.44, p=NR	med PFS = 0.83 + 0.46 * CR
Zhu 2017[81]	CR	PFS	NHL (mantle cell)	NR	Chemo, immune or targeted	NR	NR			WLR R2 (CR vs. med PFS)	R2=0.39, p=NR	
CR vs. OS												
Agarwal 2017[20]	CR	OS	Acute myeloid leukemia	1st	Systemic	20†	NR			WLR adj R2 (logit CR vs. log med OS)	Adj R2=0.48, p=NR	
Pang 2018[61]	CR	OS	Gastroesophageal	1st + 2nd	Targeted	18	7892	Correlation (NR) (CR vs. med OS)	r=0.43, p=0.18			
Li 2019[49]	CR	OS	Lung (NSCLC)	1st + 2nd	Immune checkpoint inhibitors	5*	4103*	Pearson (CR vs. med OS)	r=0.19, p=0.75	LR (CR vs. med OS)	R2=0.04, p=NR	
Nickolich 2014[59]	CR	OS	Lung (SCLC)	- 1st + 2nd + maintenance - Limited or extensive	Various	66†	8471†	Pearson (CR vs. med OS)	r=0.62, p<0.0001			
Nickolich 2014[59]	CR	OS	Lung (SCLC)	- 1st + 2nd + maintenance - Limited disease	Various	66†	8471†	Pearson (CR vs. med OS)	r=-0.04, p=0.863			
Nickolich 2014[59]	CR	OS	Lung (SCLC)	- 1st + 2nd + maintenance - Extensive disease	Various	66†	8471†	Pearson (CR vs. med OS)	r=0.19, p=0.295			

Ref	SO	FO	Cancer	Line Sub-groups	Treatment	N stds	N pts	Absolute correlation methods	Correlation coefficient (95% CI), p-value	Absolute regression Methods	Regression R2 (95% CI), p-value	Linear regression equation
PR (or VGPR or CR) vs. PFS												
Nickolich 2014[59]	PR	PFS	Lung (SCLC)	- 1st + 2nd + maintenance - Limited or extensive	Various	66†	8471†	Pearson (PR vs. med PFS)	r=0.35, p=0.019			
Nickolich 2014[59]	PR	PFS	Lung (SCLC)	- 1st + 2nd + maintenance - Limited disease	Various	66†	8471†	Pearson (PR vs. med PFS)	r=0.70, p=0.011			
Nickolich 2014[59]	PR	PFS	Lung (SCLC)	- 1st + 2nd + maintenance - Extensive disease	Various	66†	8471†	Pearson (PR vs. med PFS)	r=0.49, p=0.035			
Mangal 2018[55] (myeloma)	VGPR or CR	PFS	Multiple myeloma	2nd +	Various	79†	13322†			WLR adj R2 (VGPR or CR vs. med PFS)	Adj R2=0.64, p=NR	
PR vs. OS												
Nickolich 2014[59]	PR	OS	Lung (SCLC)	- 1st + 2nd + maintenance - Limited or extensive	Various	66†	8471†	Pearson (PR vs. med OS)	r=0.29, p=0.018			
Nickolich 2014[59]	PR	OS	Lung (SCLC)	- 1st + 2nd + maintenance - Limited disease	Various	66†	8471†	Pearson (PR vs. med OS)	r=0.60, p=0.009			
Nickolich 2014[59]	PR	OS	Lung (SCLC)	- 1st + 2nd + maintenance - Extensive disease	Various	66†	8471†	Pearson (PR vs. med OS)	r=0.66, p=0.0002			

*Calculated from reported data. †Unclear for individual subgroups.

adj, adjusted; chemo, chemotherapy; CI, confidence interval; CR, complete response; FO, final outcome; HER2, human epidermal growth factor receptor 2; log, logarithm; LR, linear regression; med, median; mo, months; NHL, non-Hodgkin's lymphoma; NR, not reported; NSCLC, non-small cell lung cancer; ORR, overall response rate (ORR=PR+CR); OS, overall survival; PFS, progression-free survival; PR, partial response; r, Pearson correlation; R2, regression coefficient of determination; r2s, squared Spearman rank correlation; rs, Spearman rank correlation; SCLC, small cell lung cancer; SO, surrogate outcome; TTP, time to progression; unwtd, unweighted; VGPR, very good partial response; wtd, weighted; WLR, weighted linear regression; WLSR, weighted least squares regression.

Supplementary Table 8: Treatment effect correlation and regression results per study

Ref	SO	FO	Cancer	Line Sub-groups	Treatment	N stds	N pts	Treatment effect correlation methods	Correlation coefficient (95% CI), p-value	Treatment effect regression methods	Regression R2 (95% CI), p-value	Linear regression equation	STE	IQWiG	BSES2
ORR vs. PFS															
Burzykowski 2008[24]	ORR	PFS	Breast	1st	Chemo	11	3953	Spearman via LR with Plackett copula (logOR ORR vs. logHR PFS)	rs=0.96 (0.73 to 1.19), p=NR	LR		logHR PFS = 0.10 + 0.50 * logOR ORR	NR	Medium+	NE
Ciani 2015[26] Elia 2018[33]	ORR	PFS	Colorectal	All	Systemic	33	NR			LR: Adj R2 (logOR ORR vs. logHR PFS)	Adj R2=0.61 (0.27 to 0.87), p=NR	logHR PFS = -0.05 - 0.32 * logOR ORR	NR	Medium	NE
Ciani 2015[26] Elia 2018[33]	ORR	PFS	Colorectal	- All - No crossover	Systemic	7	NR			LR: Adj R2 (logOR ORR vs. logHR PFS)	Adj R2=0.63 (0.03 to 0.99), p=NR	logHR PFS = -0.05 - 0.31 * logOR ORR	NR	Medium	NE
Tsujino 2010[77]	ORR	PFS	Colorectal	NR	Targeted	7	NR			LR (unwtd) R2 (diff ORR vs. HR PFS)	R2=0.65, p=0.029	Slope -0.037	NR	Medium	NE
Blumenthal 2017[22]	ORR	PFS	Lung (NSCLC)	Various	Chemo, immune or targeted	25	20013†			WLR R2: a) OR ORR vs. HR PFS b) 6mo ratio ORR vs. HR PFS	a) R2=0.74 (0.55 to 0.88), p=NR b) R2=0.70 (0.50 to 0.84), p=NR		NR	Medium+	NE
Blumenthal 2015[21]	ORR	PFS	Lung (NSCLC)	Various	Chemo or targeted	14	12567†			WLR R2 (logOR ORR vs. logHR PFS)	R2=0.89 (0.80 to 0.98), p=NR		NR	Medium+	NE
Blumenthal 2015[21]	ORR	PFS	Lung (NSCLC)	Various	Chemo	11	11701†			WLR R2 (logOR ORR vs. logHR PFS)	R2=0.77 (0.58 to 0.96), p=NR		NR	Medium+	NE
Ito 2019[45]	ORR	PFS	Lung (NSCLC)	Various	Immune (PD-(L)1)	6	3752†	a) Pearson, wtd b) Spearman, wtd (OR ORR vs. HR PFS)	a) r= -0.87, p<0.0001 b) rs= -0.97, p<0.0001	WLR R2 (OR ORR vs. HR PFS)	R2=0.76, p=0.011		NR	Medium+	Fair
Ito 2019[45]	ORR	PFS	Lung (NSCLC)	- Various - High PD-L1 expression	Immune checkpoint inhibitors (PD-(L)1)	7	1381	a) Pearson, wtd b) Spearman, wtd (OR ORR vs. HR PFS)	a) r=0.67, p<0.0001 b) rs=0.56, p<0.0001	WLR R2 (OR ORR vs. HR PFS)	R2=0.45, p=0.101		NR	Low	Good
Ritchie 2018[65]	ORR	PFS	Lung (NSCLC)	All	Immune checkpoint inhibitors (PD-(L)1 or CTLA4)	8	NR	Correlation (NR), wtd (OR ORR vs. HR PFS)	r=0.74 (0.38 to 1.08), p=NR				NR	Medium	Good

Ref	SO	FO	Cancer	Line Sub-groups	Treatment	N stds	N pts	Treatment effect correlation methods	Correlation coefficient (95% CI), p-value	Treatment effect regression methods	Regression R2 (95% CI), p-value	Linear regression equation	STE	IQWiG	BSES2
Roviello 2017[67]	ORR	PFS	Lung (NSCLC)	Various	Immune checkpoint inhibitors	7*	3369*			WLR R2 (logOR ORR vs. logHR PFS)	R2=0.42 (0.003 to 0.85), p=0.06		NR	Low	NE
Tsujino 2010[77]	ORR	PFS	Lung (NSCLC)	NR	Targeted	6	NR			LR (unwtd) R2 (diff ORR vs. HR PFS)	R2=0.94, p=0.002	Slope -0.015	NR	Medium+	NE
Colloca & Venturino 2017[27]	ORR	PFS	Ovarian	1st	Chemo	29	NR	Spearman (diff ORR vs. diff med PFS)	rs=0.64, p<0.001	LR R2 (log RR ORR vs. log HR PFS)	R2=0.28, p=0.005		NR	Low	NE
Colloca & Venturino 2017[27]	ORR	PFS	Ovarian	- 1st - Published 1990-2002	Chemo	15	NR	Spearman (diff ORR vs. diff med PFS)	rs=0.64, p=0.018	LR R2 (log RR ORR vs. log HR PFS)	R2=0.32, p=0.046		NR	Low	NE
Colloca & Venturino 2017[27]	ORR	PFS	Ovarian	- 1st - Published 2003-2016	Chemo	16	NR	Spearman (diff ORR vs. diff med PFS)	rs=0.58, p=0.019	LR R2 (log RR ORR vs. log HR PFS)	R2=0.53, p=0.003		NR	Medium	NE
Siddiqui 2017[72]	ORR	PFS	Ovarian	2nd +	Chemo	39†	9223†	Pearson, wtd (OR ORR vs. HR PFS)	r=0.42, p=NR				NR	Low	Poor
Colloca 2016a[28]	ORR	PFS	Pancreatic	1st	Gemcitabine + chemo or targeted	33*	NR	Spearman (diff ORR vs. diff med PFS)	rs=0.34, p=NR				NR	Low	NE
Colloca 2016a[28]	ORR	PFS	Pancreatic	1st	Gemcitabine + targeted	14*	NR	Spearman (diff ORR vs. diff med PFS)	rs=0.25, p=NR				NR	Low	NE
Ritchie 2018[65]	ORR	PFS	Various solid tumours	All	Immune checkpoint inhibitors (PD-(L)1 or CTLA4)	20†	10828†	Correlation (NR), wtd (OR ORR vs. HR PFS)	r=0.63 (0.35 to 0.89), p=NR				NR	Medium	Poor
Roviello 2017[67]	ORR	PFS	Various solid tumours	Various	Immune checkpoint inhibitors	17†	8994†			WLR R2 (logOR ORR vs. logHR PFS)	R2=0.32 (0.02 to 0.76), p=0.01	logHR PFS = -0.1281 - 0.2384 * logOR ORR	NR	Low	NE
Roviello 2017[67]	ORR	PFS	Various solid tumours	Various	Immune checkpoint inhibitors (CTLA-4)	17†	8994†			WLR R2 (logOR ORR vs. logHR PFS)	R2=0.67 (0.02 to 1.00), p=0.05		NR	Medium	NE
Roviello 2017[67]	ORR	PFS	Various solid tumours	Various	Immune checkpoint inhibitors (PD-(L)1)	17†	8994†			WLR R2 (logOR ORR vs. logHR PFS)	R2=0.25 (0.02 to 1.00), p=0.08		NR	Low	NE

Ref	SO	FO	Cancer	Line Sub-groups	Treatment	N stds	N pts	Treatment effect correlation methods	Correlation coefficient (95% CI), p-value	Treatment effect regression methods	Regression R2 (95% CI), p-value	Linear regression equation	STE	IQWiG	BSES2
Tsujino 2010[77]	ORR	PFS	Various solid tumours	NR	Targeted	17	NR			LR (unwtd) R2 (diff ORR vs. HR PFS)	R2=0.50, p=0.001	Slope -0.022	15%	Medium	NE
Wilkerson+ Fojo 2009[79]	ORR	PFS	Various solid tumours	NR	NR	66†	NR			LR (unwtd R2): a) diff ORR vs. HR PFS b) diff ORR vs. diff med PFS	a) R2=0.45, p<0.0001 b) R2=0.62, p<0.0001		NR	Medium	NE
ORR vs. OS															
Moriwaki 2016[56]	ORR	OS	Biliary tract	1st	Chemo	17†	2040			WLR R2 (ratio ORR vs. log ratio med OS)	R2=0.29 (0.01 to 0.65), p=0.021	log ratio med OS = 0.013 + 0.282 * ratio ORR	NR	Low	NE
Moriwaki 2016[56]	ORR	OS	Biliary tract	1st	Chemo (gemcitabine)	14†	1880			WLR R2 (ratio ORR vs. log ratio med OS)	R2=0.39 (0.02 to 0.75), p=0.013	log ratio med OS = 0.020 + 0.268 * ratio ORR	NR	Low	NE
Moriwaki 2016[56]	ORR	OS	Biliary tract	1st	Targeted	6†	953			WLR R2 (ratio ORR vs. log ratio med OS)	R2=0.43 (0.03 to 0.89), p=0.090	log ratio med OS = 0.119 + 0.155 * ratio ORR	NR	Low	NE
Bruzzi 2005[23]	ORR	OS	Breast	All	Chemo	10	2126			WLR R2: a) logOR ORR vs. logHR OS b) diff ORR vs. diff med OS	a) R2=0.10 (0.00 to 0.43), p=NR b) R2=0.20 (0 to 0.65), p=NR		NR	Low	NE
Burzykowski 2008[24]	ORR	OS	Breast	1st	Chemo	11	3953	Spearman via LR with Plackett copula (logOR ORR vs. logHR OS)	rs=0.57 (-0.31 to 1.44), p=NR				NR	Medium	NE
Hackshaw 2005[36]	ORR	OS	Breast	1st	Chemo	42*	9163			WLR R2 (logOR ORR vs. logHR OS)	R2=0.34, p<0.0001	logHR OS = -0.0081 + 0.28 * logOR ORR Slope 0.28	NR	Low	NE
Hackshaw 2005[36]	ORR	OS	Breast	- 1st - Recruited pre-1990	Chemo	26*	5244*			WLR R2 (logOR ORR vs. logHR OS)	R2=0.26, p=0.004	Slope 0.28	NR	Low	NE

Ref	SO	FO	Cancer	Line Sub-groups	Treatment	N stds	N pts	Treatment effect correlation methods	Correlation coefficient (95% CI), p-value	Treatment effect regression methods	Regression R2 (95% CI), p-value	Linear regression equation	STE	IQWiG	BSES2
Hackshaw 2005[36]	ORR	OS	Breast	- 1st - Recruited 1990 or after	Chemo	16*	3919*			WLR R2 (logOR ORR vs. logHR OS)	R2=0.41, p=0.005	Slope 0.24	NR	Low	NE
Buyse 2000[25]	ORR	OS	Colorectal	1st	Chemo	25	3791			WLR R2 (logOR ORR vs. logHR OS)	R2=0.38 (0.09 to 0.68), p=NR		NR	Low	NE
Ciani 2015[26] Elia 2018[33]	ORR	OS	Colorectal	All	Systemic	32	NR	Spearman (logOR ORR vs. logOR OS)	rs=0.53, p<0.01	a) WLSR R2 (logOR ORR vs. logOR OS) (timepoint NR) b) Adj R2 (logOR ORR vs. logHR OS)	a) R2=0.06 (0.01 to 0.29), p=NR b) Adj R2=0.33 (0.00 to 0.91), p=NR	logHR OS = -0.03 - 0.05 * logOR ORR	0.28	Low	NE
Ciani 2015[26] Elia 2018[33]	ORR	OS	Colorectal	- All - No crossover	Systemic	7	NR			LR: Adj R2 (logOR ORR vs. logHR OS)	Adj R2=0.40 (0.00 to 0.96), p=NR	logHR OS = -0.04 - 0.10 * logOR ORR	NR	Low	NE
Colloca 2016b[29]	ORR	OS	Colorectal	1st	Bevacizumab + chemo	11	NR	Spearman (diff ORR vs. diff med OS)	rs=0.82, p<0.001	LR R2 (diff ORR vs. diff med OS)	R2=0.58, p=0.002		NR	Medium	NE
Cremolini 2017[31]	ORR	OS	Colorectal	2nd	Targeted	20*	7571	Pearson (via WLR): a) rr ORR vs. HR OS b) diff ORR vs. diff med OS	a) r=0.17, p=0.476 b) r=0.35, p=0.092	WLR R2: a) rr ORR vs. HR OS b) diff ORR vs. diff med OS	b) R2=0.03, p=0.476 b) R2=0.12, p=0.092	a) Slope -0.029 b) Slope 0.071	NR	Low	NE
Cremolini 2017[31]	ORR	OS	Colorectal	2nd	Targeted, anti-angiogenic	13*	NR	Pearson (via WLR): a) rr ORR vs. HR OS b) diff ORR vs. diff med OS	a) r=0.36, p=0.249 b) r=0.52, p=0.038	WLR R2: a) rr ORR vs. HR OS b) diff ORR vs. diff med OS	b) R2=0.13, p=0.249 b) R2=0.27, p=0.038	a) Slope -0.113 b) Slope 0.133	NR	Low	NE
Cremolini 2017[31]	ORR	OS	Colorectal	2nd	Targeted, not anti-angiogenic	7*	NR	Pearson (via WLR): a) rr ORR vs. HR OS b) diff ORR vs. diff med OS	a) r=0.44, p=0.274 b) r=0.63, p=0.068	WLR R2: a) rr ORR vs. HR OS b) diff ORR vs. diff med OS	b) R2=0.20, p=0.274 b) R2=0.40, p=0.068	a) Slope -0.064 b) Slope 0.143	NR	Low	NE
Johnson 2006[46]	ORR	OS	Colorectal	1st	Chemo	146†	35337†			WLSR R2 (diff ORR vs. diff med OS)	R2=0.10, p<0.0001	Diff med OS = 0.340 + 0.096 * diff ORR	NR	Low	NE

Ref	SO	FO	Cancer	Line Sub-groups	Treatment	N stds	N pts	Treatment effect correlation methods	Correlation coefficient (95% CI), p-value	Treatment effect regression methods	Regression R2 (95% CI), p-value	Linear regression equation	STE	IQWiG	BSES2
Sidhu 2013[73]	ORR	OS	Colorectal	1st (most)	Chemo +/- targeted	24†	20438†	Correlation (NR): a) OR ORR vs. HR OS b) Diff ORR vs. HR OS c) Ratio ORR vs. HR OS	a) r=0.62 (0.37 to 0.79), p=NR b) r=0.64 (0.39 to 0.79), p=NR c) r=0.52 (0.23 to 0.72), p=NR	LR (unwtd) R2: a) OR ORR vs. HR OS b) Diff ORR vs. HR OS c) Ratio ORR vs. HR OS	a) R2=0.39 (0.13 to 0.62), p=NR b) R2=0.41 (0.15 to 0.63), p=NR c) R2=0.27 (0.05 to 0.52), p=NR		NR	Medium	NE
Sidhu 2013[73]	ORR	OS	Colorectal	1st (most)	Targeted + chemo	13	12060*	Correlation (NR): a) OR ORR vs. HR OS b) Diff ORR vs. HR OS c) Ratio ORR vs. HR OS	a) r=0.50 (0.05 to 0.75), p=NR b) r=0.58 (0.19 to 0.80), p=NR c) r=0.42 (0.00 to 0.71), p=NR	LR (unwtd) R2: a) OR ORR vs. HR OS b) Diff ORR vs. HR OS c) Ratio ORR vs. HR OS	a) R2=0.25 (0.00 to 0.57), p=NR b) R2=0.33 (0.04 to 0.64), p=NR c) R2=0.18 (0.00 to 0.51), p=NR		NR	Medium	NE
Sidhu 2013[73]	ORR	OS	Colorectal	1st (most)	Targeted (anti-EGFR)	9	7792*	Correlation (NR): a) OR ORR vs. HR OS b) Diff ORR vs. HR OS c) Ratio ORR vs. HR OS	a) r=0.67 (0.27 to 0.86), p=NR b) r=0.72 (0.35 to 0.88), p=NR c) r=0.52 (0.00 to 0.79), p=NR	LR (unwtd) R2: a) OR ORR vs. HR OS b) Diff ORR vs. HR OS c) Ratio ORR vs. HR OS	a) R2=0.45 (0.07 to 0.74), p=NR b) R2=0.52 (0.12 to 0.78), p=NR c) R2=0.27 (0.00 to 0.62), p=NR		NR	Medium	NE
Sidhu 2013[73]	ORR	OS	Colorectal	1st (most)	Targeted (anti-EGFR), KRAS non-mutant	6*	4916*	Correlation (NR): a) OR ORR vs. HR OS b) Diff ORR vs. HR OS c) Ratio ORR vs. HR OS	a) r=0.68 (0.07 to 0.89), p=NR b) r=0.81 (0.38 to 0.94), p=NR c) r=0.48 (0.00 to 0.82), p=NR	LR (unwtd) R2: a) OR ORR vs. HR OS b) Diff ORR vs. HR OS c) Ratio ORR vs. HR OS	a) R2=0.46 (0.01 to 0.80), p=NR b) R2=0.65 (0.15 to 0.88), p=NR c) R2=0.23 (0.00 to 0.67), p=NR		NR	Medium	NE
Tang 2007[75]	ORR	OS	Colorectal	1st	Chemo	39	18668	Spearman (diff ORR vs. diff med OS)	rs=0.39 (0.08 to 0.63), p=0.015				NR	Low	Poor
Tsujino 2010[77]	ORR	OS	Colorectal	NR	Targeted	7	NR			LR (unwtd) R2 (diff ORR vs. HR OS)	R2=0.51, p=0.072	Slope -0.029	NR	Medium	NE
Blumenthal 2017[22]	ORR	OS	Lung (NSCLC)	Various	Chemo, immune or targeted	25	20013†			WLR R2: a) OR ORR vs. HR OS b) 6mo ratio ORR vs. HR OS	a) R2=0.04 (0.0002 to 0.28), p=NR b) R2=0.05 (0.0001 to 0.31), p=NR		NR	Low	NE

Ref	SO	FO	Cancer	Line Sub-groups	Treatment	N stds	N pts	Treatment effect correlation methods	Correlation coefficient (95% CI), p-value	Treatment effect regression methods	Regression R2 (95% CI), p-value	Linear regression equation	STE	IQWiG	BSES2
Blumenthal 2015[21]	ORR	OS	Lung (NSCLC)	Various	Chemo or targeted	14	12567†			WLR R2 (logOR ORR vs. logHR OS)	R2=0.09 (0 to 0.33), p=NR		NR	Low	NE
Blumenthal 2015[21]	ORR	OS	Lung (NSCLC)	Various	Chemo	11	11701†			WLR R2 (logOR ORR vs. logHR OS)	R2=0.44 (0.08 to 0.80), p=NR		NR	Low	NE
Hashim 2018[39]	ORR	OS	Lung (NSCLC)	2nd +	Various	140	41725	Correlation (NR) via WLR: a) diff ORR vs. logHR OS b) diff ORR vs. diff med OS	a) r=0.17 (0.00 to 0.38), p=NR b) r=0.18 (0.02 to 0.34), p=0.032				NA	Low	NE
Hashim 2018[39]	ORR	OS	Lung (NSCLC)	- 2nd + - Phase III	Various	59	32348	Correlation (NR) via WLR: a) diff ORR vs. logHR OS b) diff ORR vs. diff med OS	a) r=0.37 (0.09 to 0.60), p=NR b) r=0.13 (0.00 to 0.38), p=0.32				NA	Low	NE
Hashim 2018[39]	ORR	OS	Lung (NSCLC)	- 2nd + - Phase III excl per-protocol crossover	Various	54	30654	Correlation (NR) via WLR: a) diff ORR vs. logHR OS b) diff ORR vs. diff med OS	a) r=0.40 (0.10 to 0.63), p=NR b) r=0.36 (0.10 to 0.57), p=0.0074				NA	Low	NE
Hashim 2018[39]	ORR	OS	Lung (NSCLC)	- 2nd + - Phase III excl any crossover	Various	38	22574	Correlation (NR) via WLR: a) diff ORR vs. logHR OS b) diff ORR vs. diff med OS	a) r=0.52 (0.18 to 0.75), p=NR b) r=0.45 (0.15 to 0.67), p=0.0051				a) 55% b) NA	Medium	NE
Hashim 2018[39]	ORR	OS	Lung (NSCLC)	- 2nd + - Phase III excl crossover or unbalanced post-progression treatments	Various	18	13349	Correlation (NR) via WLR: a) diff ORR vs. logHR OS b) diff ORR vs. diff med OS	a) r=0.16 (0.00 to 0.60), p=NR b) r=0.53 (0.08 to 0.80), p=0.024				a) NA b) 41%	Low	NE
Hotta 2015[40]	ORR	OS	Lung (NSCLC)	Various	Targeted	18	7633†			WLR R2 (OR ORR vs. HR OS)	R2=0.10, p=NR		NR	Low	NE

Ref	SO	FO	Cancer	Line Sub-groups	Treatment	N stds	N pts	Treatment effect correlation methods	Correlation coefficient (95% CI), p-value	Treatment effect regression methods	Regression R2 (95% CI), p-value	Linear regression equation	STE	IQWiG	BSES2
Hotta 2015[40]	ORR	OS	Lung (NSCLC)	- Various - Molecularly selected	Targeted	8	NR			WLR R2 (OR ORR vs. HR OS)	R2=0.04, p=NR		NR	Low	NE
Hotta 2015[40]	ORR	OS	Lung (NSCLC)	- Various - Non-molecularly selected	Targeted	10	NR			WLR R2 (OR ORR vs. HR OS)	R2=0.43, p=NR		NR	Low	NE
Ito 2019[45]	ORR	OS	Lung (NSCLC)	Various	Immune checkpoint inhibitors (PD-(L)1)	6	3752†	a) Pearson, wtd b) Spearman, wtd (OR ORR vs. HR OS)	a) r= -0.75, p<0.0001 b) rs= -0.96, p<0.0001	WLR R2 (OR ORR vs. HR OS)	R2=0.57, p=0.051		NR	Medium	Poor
Ito 2019[45]	ORR	OS	Lung (NSCLC)	- Various - High PD-L1 expression	Immune checkpoint inhibitors (PD-(L)1)	7	1381	a) Pearson, wtd b) Spearman, wtd (OR ORR vs. HR OS)	a) r= -0.50, p<0.0001 b) rs= -0.21, p<0.0001	WLR R2 (OR ORR vs. HR OS)	R2=0.25, p=0.253		NR	Low	Fair
Johnson 2006[46]	ORR	OS	Lung (NSCLC)	1st	Chemo	191†	44125†			WLSR R2 (diff ORR vs. diff med OS)	R2=0.16, p<0.0001	Diff med OS = -0.048 + 0.090 * diff ORR	NR	Low	NE
Nakashima 2016[58]	ORR	OS	Lung (NSCLC)	1st	Chemo	44	22709	Spearman, wtd (lnOR ORR vs. HR OS)	rs=0.57, p=NR	WLSR adj R2 (lnOR ORR vs. lnHR OS)	Adj R2=0.35, p=NR	lnHR OS = -0.023 -0.133 x lnOR ORR	NR	Low	NE
Ritchie 2018[65]	ORR	OS	Lung (NSCLC)	All	Immune checkpoint inhibitors (PD-(L)1 or CTLA4)	8	NR	Correlation (NR), wtd (OR ORR vs. HR OS)	r=0.68 (0.08 to 1.10), p=NR				NR	Low	Good
Roviello 2017[67]	ORR	OS	Lung (NSCLC)	Various	Immune checkpoint inhibitors	7*	3369*			WLR R2 (logOR ORR vs. logHR OS)	R2=0.0007 (0.09 to 0.91), p=0.94		NR	Low	NE
Tsujino 2010[77]	ORR	OS	Lung (NSCLC)	NR	Targeted	5	NR			LR (unwtd) R2 (diff ORR vs. HR OS)	R2=0.84, p=0.030	Slope -0.011	NR	Medium+	NE
Foster 2011[34]	ORR	OS	Lung (SCLC)	1st	Chemo	3 (32 centres)	596†	Spearman (logOR ORR vs. logHR OS)	rs=0.52, p=NR	WLSR R2 (logOR ORR vs. logHR OS)	R2=0.21, p=NR		NR	Low	NE
Hotta 2009[41]	ORR	OS	Lung (SCLC)	1st	Chemo	48	8779			WLR R2 (rr ORR vs. diff med OS)	R2=0.33, p=NR	Diff med OS = 0.00 + 0.06 * rr ORR	NR	Low	NE

Ref	SO	FO	Cancer	Line Sub-groups	Treatment	N stds	N pts	Treatment effect correlation methods	Correlation coefficient (95% CI), p-value	Treatment effect regression methods	Regression R2 (95% CI), p-value	Linear regression equation	STE	IQWiG	BSES2
Hotta 2009[41]	ORR	OS	Lung (SCLC)	- 1st - Clear criteria	Chemo	43 comp	NR			WLR R2 (rr ORR vs. diff med OS)	R2=0.19, p=NR		NR	Low	NE
Hotta 2009[41]	ORR	OS	Lung (SCLC)	- 1st - WHO criteria	Chemo	23 comp	NR			WLR R2 (rr ORR vs. diff med OS)	R2=0.13, p=NR		NR	Low	NE
Hotta 2009[41]	ORR	OS	Lung (SCLC)	- 1st - Non-WHO criteria	Chemo	20 comp	NR			WLR R2 (rr ORR vs. diff med OS)	R2=0.28, p=NR		NR	Low	NE
Hotta 2009[41]	ORR	OS	Lung (SCLC)	- 1st - Published 1990-1996	Chemo	26 comp	NR			WLR R2 (rr ORR vs. diff med OS)	R2=0.23, p=NR	Diff med OS = 0.00 + 0.04 * rr ORR	NR	Low	NE
Hotta 2009[41]	ORR	OS	Lung (SCLC)	- 1st - Published 1997-2008	Chemo	26 comp	NR			WLR R2 (rr ORR vs. diff med OS)	R2=0.47, p=NR	Diff med OS = 0.00 + 0.09 * rr ORR	NR	Low	NE
Colloca & Venturino 2017[27]	ORR	OS	Ovarian	1st	Chemo	27	NR	Spearman (diff ORR vs. diff med OS)	rs=0.41, p=0.035	LR R2 (log RR ORR vs. log HR OS)	R2=0.12, p=0.073		NR	Low	NE
Colloca & Venturino 2017[27]	ORR	OS	Ovarian	- 1st - Published 1990-2002	Chemo	13	NR	Spearman (diff ORR vs. diff med OS)	rs=0.65, p=0.016	LR R2 (log RR ORR vs. log HR OS)	R2=0.15, p=0.199		NR	Low	NE
Colloca & Venturino 2017[27]	ORR	OS	Ovarian	- 1st - Published 2003-2016	Chemo	14	NR	Spearman (diff ORR vs. diff med OS)	rs= -0.02, p=0.940	LR R2 (log RR ORR vs. log HR OS)	R2=0.34, p=0.027		NR	Low	NE
Siddiqui 2017[72]	ORR	OS	Ovarian	2nd +	Chemo	31†	9223†						NR	NE	NE
Colloca 2016a[28]	ORR	OS	Pancreatic	1st	Gemcitabine + chemo or targeted	36*	NR	Spearman (diff ORR vs. diff med OS)	rs=0.29, p=0.067				NR	Low	NE
Colloca 2016a[28]	ORR	OS	Pancreatic	1st	Gemcitabine + chemo	22*	NR	Spearman (diff ORR vs. diff med OS)	rs=0.23, p=0.250	LR R2 (logRR ORR vs. logHR OS)	R2=0.15, p=NR		NR	Low	NE
Colloca 2016a[28]	ORR	OS	Pancreatic	1st	Gemcitabine + targeted	14*	NR	Spearman (diff ORR vs. diff med OS)	rs=0.55, p=0.035	LR R2 (logRR ORR vs. logHR OS)	R2=0.28, p=NR		NR	Low	NE
Hamada 2016[37]	ORR	OS	Pancreatic	1st	Chemo	36	15906†	Spearman via WLSR (logOR ORR vs. logHR OS)	rs= -0.16 (-0.27 to -0.05), p=0.007	WLSR adj R2 (logOR ORR vs. logHR OS)	Adj R2=0.30, p=0.007		NR	Low	Poor

Ref	SO	FO	Cancer	Line Sub-groups	Treatment	N stds	N pts	Treatment effect correlation methods	Correlation coefficient (95% CI), p-value	Treatment effect regression methods	Regression R2 (95% CI), p-value	Linear regression equation	STE	IQWiG	BSES2
Makris 2017[53]	ORR	OS	Pancreatic (adenocarcinoma)	1st	Chemo (gemcitabine)	22*	10379*	Pearson (log HR OS vs. log OR ORR): a) wtd by sample size b) fixed effect c) random effects	a) r=0.27 (-0.14 to 0.60), p=0.20 b) r=0.52 (0.16 to 0.76), p=0.007 c) r=0.45 (0.07 to 0.72), p=0.02				NR	Low	NE
Makris 2017[53]	ORR	OS	Pancreatic (adenocarcinoma)	- 1st - No crossover	Chemo (gemcitabine)	22*	10379*	Pearson (log HR OS vs. log OR ORR): a) wtd by sample size b) fixed effect c) random effects	a) r= -0.10 (-0.56 to 0.40), p=0.70 b) r=0.16 (-0.34 to 0.60), p=0.53 c) r=0.21 (-0.30 to 0.62), p=0.43				NR	Low	NE
Makris 2017[53]	ORR	OS	Pancreatic (adenocarcinoma)	- 1st - Crossover <50%	Chemo (gemcitabine)	22*	10379*	Pearson (log HR OS vs. log OR ORR): a) wtd by sample size b) fixed effect c) random effects	a) r=0.26 (-0.18 to 0.62), p=0.24 b) r=0.53 (0.15 to 0.78), p=0.009 c) r=0.45 (0.03 to 0.73), p=0.03				NR	Low	NE
Colloca 2016c[30]	ORR	OS	Prostate	1st + 2nd	Chemo, hormonal + targeted	17	NR	Pearson (diff ORR vs. diff med OS)	r=0.38, p=0.132	LR R2 (log RR ORR vs. log HR OS)	R2=0.007, p=0.789		NR	Low	NE
Colloca 2016c[30]	ORR	OS	Prostate	- 1st + 2nd - Published 1995-2004	Chemo, hormonal + targeted	5	NR	Pearson (diff ORR vs. diff med OS)	r=0.35, p=0.560	LR R2 (log RR ORR vs. log HR OS)	R2=0.53, p=0.275		NR	Medium	NE
Colloca 2016c[30]	ORR	OS	Prostate	- 1st + 2nd - Published 2005-2014	Chemo, hormonal + targeted	12	NR	Pearson (diff ORR vs. diff med OS)	r=0.41, p=0.185	LR R2 (log RR ORR vs. log HR OS)	R2=0.02, p=0.690		NR	Low	NE
Delea 2012[32]	ORR	OS	Renal cell	NR	Cytokine or targeted	25*	10943†	Pearson, wtd (ln(rr) ORR vs. -lnHR OS)	r=0.78, p<0.0001	WLSR adj R2 (ln rr ORR vs. -lnHR OS)	Adj R2=0.59, p<0.0001	-lnHR OS = -0.11 + 0.30 * lnrr ORR	NR	Medium	NE
Petrelli 2013[63]	ORR	OS	Renal cell	1st	Targeted	6†	3188†	a) Pearson, wtd b) Spearman, wtd (diff med OS vs. diff ORR)	a) r =0.52, p<0.0001 b) rs = 0.49, p<0.0001	LR	R2=0.27, p=NR		NR	Low	Fair
Tanaka 2019[74]	ORR	OS	Soft tissue sarcoma	1st	Chemo	27†	6156†	Kendall's Tau (logOR ORR vs. logHR OS)	τ=0.41, p=NR	Regression (NR) R2 (logOR ORR vs. logHR OS)	R2=0.28 (0.02 to 0.54), p=NR		NR	Low	NE

Ref	SO	FO	Cancer	Line Sub-groups	Treatment	N stds	N pts	Treatment effect correlation methods	Correlation coefficient (95% CI), p-value	Treatment effect regression methods	Regression R2 (95% CI), p-value	Linear regression equation	STE	IQWiG	BSES2
Zer 2016[80]	ORR	OS	Soft tissue sarcoma	All	Systemic	52†	9762†	Correlation (NR) via WLR (OR ORR vs. HR OS)	r=0.51, p=NR				NR	Low	NE
Kaufman 2018[47]	ORR	OS	Various solid tumours	Various	Immune checkpoint inhibitors + chemo	27†	10300†			WLR adj R2 (OR ORR vs. HR OS)	Adj R2= -0.07, p=0.866		NR	NE	NE
Kaufman 2018[47]	ORR	OS	Various solid tumours	Various	Immune checkpoint inhibitors alone	NR	NR			WLR adj R2 (OR ORR vs. HR OS)	Adj R2= -0.08, p=0.799		NR	NE	NE
Mushti 2018[57]	ORR	OS	Various solid tumours	NR	Immune checkpoint inhibitors (PD-(L)1)	13	6722			WLR R2 (OR ORR vs. HR OS)	R2=0.13, p=NR		NR	Low	NE
Nie 2019[60]	ORR	OS	Various solid tumours	Various	Immune checkpoint inhibitors (PD-(L)1)	43†	15088†			WLR R2 (lnOR ORR vs. lnHR OS)	R2=0.10, p=0.053		NR	Low	Poor
Ritchie 2018[65]	ORR	OS	Various solid tumours	All	Immune checkpoint inhibitors (PD-(L)1 or CTLA4)	20†	10828†	Correlation (NR), wtd (OR ORR vs. HR OS)	r=0.57 (0.23 to 0.89), p=NR				NR	Low	Poor
Roviello 2017[67]	ORR	OS	Various solid tumours	Various	Immune checkpoint inhibitors	17†	8994†			WLR R2 (logOR ORR vs. logHR OS)	R2=0.47 (0.03 to 0.77), p=0.001	logHR OS = -0.1329 - 0.2575 * logOR ORR	NR	Low	NE
Roviello 2017[67]	ORR	OS	Various solid tumours	Various	Immune checkpoint inhibitors (CTLA-4)	17†	8994†			WLR R2 (logOR ORR vs. logHR OS)	R2=0.00 (0.00 to 0.97), p=0.96		NR	Low	NE
Roviello 2017[67]	ORR	OS	Various solid tumours	Various	Immune checkpoint inhibitors (PD-(L)1)	17†	8994†			WLR R2 (logOR ORR vs. logHR OS)	R2=0.18 (0.00 to 0.97), p=0.17		NR	Low	NE
Tsujino 2010[77]	ORR	OS	Various solid tumours	NR	Targeted	18	NR			LR (unwtd) R2 (diff ORR vs. HR OS)	R2=0.47, p=0.002	Slope -0.016	21%	Low	NE

Ref	SO	FO	Cancer	Line Sub-groups	Treatment	N stds	N pts	Treatment effect correlation methods	Correlation coefficient (95% CI), p-value	Treatment effect regression methods	Regression R2 (95% CI), p-value	Linear regression equation	STE	IQWiG	BSES2
Wilkerson+ Fojo 2009[79]	ORR	OS	Various solid tumours	NR	NR	66†	NR			LR (unwtd R2): a) diff ORR vs. HR OS b) diff ORR vs. diff med OS	a) R2=0.37, p<0.0001 b) R2=0.34, p<0.0001		NR	Low	NE
CR vs. PFS															
Lee 2011[48]	CR	PFS	NHL (aggressive)	1st	Chemo	12†	NR	Spearman (diff CR vs. diff 3yr PFS)	rs=0.63 (0.21 to 0.84), p=0.005				NR	Medium	NE
Lee 2011[48]	CR	PFS	NHL (indolent)	1st	Chemo	6†	NR	Spearman (diff CR vs. diff 3yr PFS)	rs=0.41 (-0.52 to 0.88), p=0.35				NR	Medium	NE
Shi 2017[69]	CR	PFS	NHL (indolent; follicular)	1st	Chemo or immuno (induction or maintenance)	13	3837			a) WLSR R2 b) Bivariate Plackett copula model (logOR CR 30mo vs. logHR PFS)	a) R2WLS=0.88 (0.77 to 0.96), p=NR b) R2Copula=0.86 (0.72 to 1.00), p=NR	logHR PFS = -0.093 - 0.636 * logOR CR 30mo	1.56	Medium+	NE
Shi 2017[69]	CR	PFS	NHL (indolent; follicular)	1st	Rituximab-based (induction or maintenance)	9	2851			a) WLSR R2 b) Bivariate Plackett copula model (logOR CR 30mo vs. logHR PFS)	a) R2WLS=0.85 (0.62 to 0.97), p=NR b) R2Copula=0.80 (0.56 to 1.00), p=NR		NR	Medium+	NE
Shi 2017[69]	CR	PFS	NHL (indolent; follicular)	1st	Non-rituximab-based (induction or maintenance)	4	986			a) WLSR R2 b) Bivariate Plackett copula model (logOR CR 30mo vs. logHR PFS)	a) R2WLS=0.91 (0.05 to 1.00), p=NR b) R2Copula=0.96 (0.90 to 1.00), p=NR		NR	Medium+	NE
Shi 2017[69]	CR	PFS	NHL (indolent; follicular)	1st	Induction	8	2207			a) WLSR R2 b) Bivariate Plackett copula model (logOR CR 30mo vs. logHR PFS)	a) R2WLS=0.89 (0.75 to 0.98), p=NR b) R2Copula=0.89 (0.74 to 1.00), p=NR		NR	Medium+	NE

Ref	SO	FO	Cancer	Line Sub-groups	Treatment	N stds	N pts	Treatment effect correlation methods	Correlation coefficient (95% CI), p-value	Treatment effect regression methods	Regression R2 (95% CI), p-value	Linear regression equation	STE	IQWiG	BSES2
Shi 2017[69]	CR	PFS	NHL (indolent; follicular)	1st	Maintenance	5	1630			wtd least squares (reported as R2WLS) and bivariate Plackett copula model (reported as R2copula), CR30 vs PFS	a) R2WLS=0.93 (0.84 to 1.00), p=NR b) R2Copula=0.89 (0.71 to 1.00), p=NR		NR	Medium+	NE
Shi 2017[69]	CR	PFS	NHL (indolent; follicular)	- 1st - High FLIPI score	Chemo or immuno (induction or maintenance)	9	1415			a) WLSR R2 b) Bivariate Plackett copula model (logOR CR 30mo vs. logHR PFS)	a) R2WLS=0.87 (0.68 to 0.98), p=NR b) R2Copula=0.73 (0.42 to 1.00), p=NR		NR	Medium+	NE
Shi 2017[69]	CR	PFS	NHL (indolent; follicular)	- 1st - Low to intermediate FLIPI score	Chemo or immuno (induction or maintenance)	10	1882			a) WLSR R2 b) Bivariate Plackett copula model (logOR CR 30mo vs. logHR PFS)	a) R2WLS=0.45 (0.02 to 0.93), p=NR b) R2Copula=0.57 (0.17 to 0.97), p=NR		NR	Low	NE
Shi 2017[69]	CR	PFS	NHL (indolent; follicular)	1st	Chemo or immuno (induction or maintenance)	11	2728			a) WLSR R2 b) Bivariate Plackett copula model (logOR CR 24mo vs. logHR PFS)	a) R2WLS=0.84 (0.63 to 0.95), p=NR b) R2Copula=0.67 (0.35 to 0.99), p=NR	logHR PFS = 0.043 - 0.726 * logOR CR24mo	NR	Medium+	NE
Shi 2017[69]	CR	PFS	NHL (indolent; follicular)	- 1st - Stage IV	Chemo or immuno (induction or maintenance)	NR	2585			a) WLSR R2 b) Bivariate Plackett copula model (logOR CR 30mo vs. logHR PFS)	a) R2WLS=0.92 (0.85 to 0.97), p=NR b) R2Copula=0.94 (0.87 to 1.00), p=NR		NR	Medium+	NE
Colloca & Venturino 2017[27]	CR	PFS	Ovarian	1st	Chemo	12	NR	Spearman (diff RR vs. diff med PFS)	rs=0.19, p=0.555				NR	Low	NE

Ref	SO	FO	Cancer	Line Sub-groups	Treatment	N stds	N pts	Treatment effect correlation methods	Correlation coefficient (95% CI), p-value	Treatment effect regression methods	Regression R2 (95% CI), p-value	Linear regression equation	STE	IQWiG	BSES2
CR vs. OS															
Hackshaw 2005[36]	CR	OS	Breast	1st	Chemo	41*	9163†			WLR R2 (logOR CR vs. logHR OS)	R2=0.12, p=0.02	logHR OS = -0.0097 + 0.13 * logOR CR Slope 0.13	NR	Low	NE
Hackshaw 2005[36]	CR	OS	Breast	- 1st - Recruited pre-1990	Chemo	26*	5244†			WLR R2 (logOR CR vs. logHR OS)	R2=0.05, p=0.24	Slope 0.09	NR	Low	NE
Hackshaw 2005[36]	CR	OS	Breast	- 1st - Recruited 1990 or after	Chemo	15*	3919†			WLR R2 (logOR CR vs. logHR OS)	R2=0.36, p=0.01	Slope 0.16	NR	Low	NE
Foster 2011[34]	CR	OS	Lung (SCLC)	1st	Chemo	3 (32 centres)	596†	Spearman (logOR CR vs. logHR OS)	rs=0.50, p=NR	WLSR R2 (logOR CR vs. logHR OS)	R2=0.48, p=NR		NR	Low	NE
Lee 2011[48]	CR	OS	NHL (aggressive)	1st	Chemo	36†	16103†	Spearman: a) diff CR vs. diff 3yr OS b) diff CR vs. diff 5yr OS	a) rs=0.58 (0.29 to 0.77), p=0.004 b) rs=0.50 (0.23 to 0.74), p=0.01				NR	Medium	NE
Lee 2011[48]	CR	OS	NHL (indolent)	1st	Chemo	15†	5128†	Spearman: a) diff CR vs. diff 3yr OS b) diff CR vs. diff 5yr OS	a) rs=0.41 (-0.10 to 0.74), p=0.098 b) rs=0.21 (-0.34 to 0.50), p=0.44				NR	Medium	NE
Colloca & Venturino 2017[27]	CR	OS	Ovarian	1st	Chemo	12	NR	Spearman (diff pCR vs. diff med OS)	rs=0.42, p=0.180				NR	Low	NE
DoR vs. OS															
Colloca 2016b[29]	DoR	OS	Colorectal	1st	Bevacizumab + chemo	5	NR	Spearman (diff med DoR vs. diff med OS)	rs=0.70, p=0.188				NR	Medium	NE
Colloca 2016a[28]	DoR	OS	Pancreatic	1st	Gemcitabine + chemo or targeted	7†	NR	Spearman (diff med DoR vs. diff med OS)	rs=0.76, p=0.049				NR	Medium	NE
Colloca 2016a[28]	DoR	OS	Pancreatic	1st	Gemcitabine + chemo	3†	NR	Spearman (diff med DoR vs. diff med OS)	rs=0.50, p=0.667				NR	Low	NE

Ref	SO	FO	Cancer	Line Sub-groups	Treatment	N stds	N pts	Treatment effect correlation methods	Correlation coefficient (95% CI), p-value	Treatment effect regression methods	Regression R2 (95% CI), p-value	Linear regression equation	STE	IQWiG	BSES2
Colloca 2016a[28]	DoR	OS	Pancreatic	1st	Gemcitabine + targeted	4†	NR	Spearman (diff med DoR vs. diff med OS)	rs=0.40, p=0.600				NR	Low	NE

*Calculated from reported data. †Unclear for individual subgroups.

adj, adjusted; BSES2, Biomarker-Surrogate Evaluation Schema criteria 2; chemo, chemotherapy; CI, confidence interval; CR, complete response; diff, difference;; DoR, duration of response; FO, final outcome; HR, hazard ratio; IQWiG, Institute of Quality and Efficiency in Health Care; ln, natural logarithm; log, logarithm; LR, linear regression; med, median; mo, months; NE, not estimable; NHL, non-Hodgkin's lymphoma; NR, not reported; NSCLC, non-small cell lung cancer; OR, odds ratio; ORR, overall response rate (ORR=PR+CR); OS, overall survival; PFS, progression-free survival; r, Pearson correlation; R2, regression coefficient of determination; rs, Spearman rank correlation; rr, relative risk; SCLC, small cell lung cancer; SO, surrogate outcome; STE, surrogate threshold effect; unwtd, unweighted; wtd, weighted; WLR, weighted linear regression; WLSR, weighted least squares regression.

12

13 **Supplementary Table 9: Influence of clinical and study factors on association between ORR and OS**

Disease and factor comparison	Absolute association (r)		Treatment effect association (R^2)	
	Range Factor A	**Range Factor B**	**Range Factor A**	**Range Factor B**
AML				
Treatment line: (A) 1st line vs (B) subsequent line	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*
Treatment type: (A) targeted vs (B) systemic	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*
Response criteria: (A) RECIST vs (B) WHO	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*
OS adjustment: (A) adjusted vs (B) unadjusted	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*
Data type: (A) Aggregate vs (B) IPD	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*
Biliary tract				
Treatment line: (A) 1st line vs (B) subsequent line	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*
Treatment type: (A) targeted vs (B) systemic	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	**0.43** [56]	**0.29 to 0.39** [56]
Response criteria: (A) RECIST vs (B) WHO	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*
OS adjustment: (A) adjusted vs (B) unadjusted	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*
Data type: (A) Aggregate vs (B) IPD	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*
Breast				
Treatment line: (A) 1st line vs (B) subsequent line	**0.61** [64]	**-0.10 to 1.00** [51]	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*
Treatment type: (A) targeted vs (B) systemic	**0.61** [64]	**-0.10 to 1.00** [51]	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*
Response criteria: (A) RECIST vs (B) WHO	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*
OS adjustment: (A) adjusted vs (B) unadjusted	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*
Data type: (A) Aggregate vs (B) IPD	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	**0.26 to 0.41** [36]	**0.10 to 0.20**[23]
Colorectal				
Treatment line: (A) 1st line vs (B) subsequent line	**0.41 to 0.59** [52,75]	**0.58** [35]	**0.10 to 0.58** [25,29,46]	**0.03 to 0.40** [31]
Treatment type: (A) targeted vs (B) systemic	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	**0.03 to 0.65** [29,31,73,77]	**0.06 to 0.40** [25,26,46]
Response criteria: (A) RECIST vs (B) WHO	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	**0.58** [29]	**0.38** [25]
OS adjustment: (A) adjusted vs (B) unadjusted	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	**0.4** [26]	**0.03 to 0.65** [25,26,29,31,46,73,77]
Data type: (A) Aggregate vs (B) IPD	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	**0.03 to 0.65** [26,29,31,46,73,77]	**0.38** [25]
Gastric and gastroesophageal				
Treatment line: (A) 1st line vs (B) subsequent line	**0.18 to 0.47** [42]	**0.38** [70]	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*

Disease and factor comparison	Absolute association (r)		Treatment effect association (R^2)	
	Range Factor A	**Range Factor B**	**Range Factor A**	**Range Factor B**
Treatment type: (A) targeted vs (B) systemic	**0.86** [61]	**0.18 to 0.47** [42,70]	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*
Response criteria: (A) RECIST vs (B) WHO	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*
OS adjustment: (A) adjusted vs (B) unadjusted	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*
Data type: (A) Aggregate vs (B) IPD	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*
Glioblastoma				
Treatment line: (A) 1st line vs (B) subsequent line	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*
Treatment type: (A) targeted vs (B) systemic	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*
Response criteria: (A) RECIST vs (B) WHO	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*
OS adjustment: (A) adjusted vs (B) unadjusted	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*
Data type: (A) Aggregate vs (B) IPD	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*
Neuroendocrine				
Treatment line: (A) 1st line vs (B) subsequent line	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*
Treatment type: (A) targeted vs (B) systemic	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*
Response criteria: (A) RECIST vs (B) WHO	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*
OS adjustment: (A) adjusted vs (B) unadjusted	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*
Data type: (A) Aggregate vs (B) IPD	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*
NSCLC				
Treatment line: (A) 1st line vs (B) subsequent line	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	**0.16 to 0.35** [46,58]	**0.03 to 0.27** [39]
Treatment type: (A) targeted vs (B) systemic	**-0.02 to 0.92** [45,49,50,65,71]	**0.41 to 0.62** [68,71]	**0.0007 to 0.84** [40,45,65,67,77]	**0.16 to 0.44** [21,46,58]
Response criteria: (A) RECIST vs (B) WHO	**0.52** [49]	**0.62** [68]	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*
OS adjustment: (A) adjusted vs (B) unadjusted	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	**0.03 to 0.27** [39]	**0.0007 to 0.84** [21,22,39,40,45,46,58,65,67,77]
Data type: (A) Aggregate vs (B) IPD	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*
Ovarian				
Treatment line: (A) 1st line vs (B) subsequent line	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*
Treatment type: (A) targeted vs (B) systemic	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*
Response criteria: (A) RECIST vs (B) WHO	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*
OS adjustment: (A) adjusted vs (B) unadjusted	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*

Disease and factor comparison	Absolute association (r)		Treatment effect association (R^2)	
	Range Factor A	**Range Factor B**	**Range Factor A**	**Range Factor B**
Data type: (A) Aggregate vs (B) IPD	**0.82** [72]	**0.56** [66]	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*
Pancreatic / adenocarcinoma				
Treatment line: (A) 1st line vs (B) subsequent line	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*
Treatment type: (A) targeted vs (B) systemic	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	**0.28** [28]	**0.01 to 0.30** [28,37,53]
Response criteria: (A) RECIST vs (B) WHO	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*
OS adjustment: (A) adjusted vs (B) unadjusted	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	**0.01 to 0.04** [53]	**0.07 to 0.30** [28,37,53]
Data type: (A) Aggregate vs (B) IPD	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*
Prostate				
Treatment line: (A) 1st line vs (B) subsequent line	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*
Treatment type: (A) targeted vs (B) systemic	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*
Response criteria: (A) RECIST vs (B) WHO	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*
OS adjustment: (A) adjusted vs (B) unadjusted	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*
Data type: (A) Aggregate vs (B) IPD	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*
Renal / renal cell				
Treatment line: (A) 1st line vs (B) subsequent line	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*
Treatment type: (A) targeted vs (B) systemic	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*
Response criteria: (A) RECIST vs (B) WHO	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*
OS adjustment: (A) adjusted vs (B) unadjusted	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*
Data type: (A) Aggregate vs (B) IPD	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*
SCLC				
Treatment line: (A) 1st line vs (B) subsequent line	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*
Treatment type: (A) targeted vs (B) systemic	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*
Response criteria: (A) RECIST vs (B) WHO	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*
OS adjustment: (A) adjusted vs (B) unadjusted	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*
Data type: (A) Aggregate vs (B) IPD	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*
Soft tissue sarcoma				
Treatment line: (A) 1st line vs (B) subsequent line	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*
Treatment type: (A) targeted vs (B) systemic	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*

Disease and factor comparison	Absolute association (r)		Treatment effect association (R^2)	
	Range Factor A	**Range Factor B**	**Range Factor A**	**Range Factor B**
Response criteria: (A) RECIST vs (B) WHO	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*
OS adjustment: (A) adjusted vs (B) unadjusted	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*
Data type: (A) Aggregate vs (B) IPD	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*
Unknown primary				
Treatment line: (A) 1st line vs (B) subsequent line	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*
Treatment type: (A) targeted vs (B) systemic	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*
Response criteria: (A) RECIST vs (B) WHO	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*
OS adjustment: (A) adjusted vs (B) unadjusted	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*
Data type: (A) Aggregate vs (B) IPD	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*
Urothelial				
Treatment line: (A) 1st line vs (B) subsequent line	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*
Treatment type: (A) targeted vs (B) systemic	***-0.12*** [18]	***-0.02 to 0.78*** [19]	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*
Response criteria: (A) RECIST vs (B) WHO	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*
OS adjustment: (A) adjusted vs (B) unadjusted	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*
Data type: (A) Aggregate vs (B) IPD	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*	*INSUFFICIENT DATA*

AML, acute myeloid leukaemia; IPD, individual patient data; NSCLC, non-small cell lung cancer; r, correlation coefficient (e.g. Pearson or Spearman); R^2, regression coefficient of determination; SCLC, small cell lung cancer.

14

Supplementary Table 10: Regression equations for absolute (individual-level) associations

Surrogate relationship	Cancer types and references	Surrogate	Final	Intercept	Slope
ORR to PFS	Colorectal[52]	ORR	Median PFS	3.20	0.10
	Lung (NSCLC)[76]	ORR	Median PFS	NR	0.07
	Ovarian[72]	ORR	Median PFS	2.59	0.12
	NHL[54]	log odds ORR	log median PFS	1.97	0.41
ORR to TTP	Gastric[42]	ORR	Median TTP	1.73	0.09
ORR to OS	Colorectal[52]	ORR	Median OS	10.45	0.09
	Lung (NSCLC)[76]	ORR	Median OS	NR	0.26
	Ovarian[72]	ORR	Median OS	9.48	0.28
	Gastric[42]	ORR	Median OS	5.89	0.08
CR to PFS	NHL[81]	CR	Median PFS	0.83	0.46
	NHL[54]	log odds CR	log median PFS	2.38	0.34

CR, complete response; NHL, non-Hodgkin's lymphoma; NSCLC, non-small cell lung cancer; ORR, overall response rate; OS, overall survival; PFS, progression-free survival; TTP, time to progression; VGPR, very good partial response.

Supplementary Table 11: Regression equations for treatment effect (trial-level) associations

Surrogate relationship	Cancer types and refs	Subgroup	Based on difference in response				Based on relative risk or odds ratio for response			
			Surrogate	Final	Intercept	Slope	Surrogate	Final	Intercept	Slope
ORR to PFS	Lung (NSCLC)[77]		Diff ORR	HR PFS	NR	-0.02				
	Colorectal[77]		Diff ORR	HR PFS	NR	-0.04				
	Various[77]		Diff ORR	HR PFS	NR	-0.02				
	Colorectal[26,33]						logOR ORR	logHR PFS	-0.05	-0.32
	Breast[24]						logOR ORR	logHR PFS	0.10	0.50
	Various (immuno)[67]						logOR ORR	logHR PFS	-0.13	-0.24
ORR to OS	Colorectal[31]	- All - Anti-angio - Non-anti-angio	Diff ORR	Diff median OS	NR	0.07 0.13 0.14				
	Colorectal[46]		Diff ORR	Diff median OS	0.34	0.10				
	Lung (NSCLC)[46]		Diff ORR	Diff median OS	-0.05	0.09				
	Colorectal[77]		Diff ORR	HR OS	NR	-0.03				
	Lung (NSCLC)[77]		Diff ORR	HR OS	NR	-0.01				
	Various[77]		Diff ORR	HR OS	NR	-0.02				
	Colorectal[26,33]	- All - No crossover					logOR ORR	logHR OS	-0.03 -0.04	-0.05 -0.10
	Breast[36]	- All - Recr. pre-1990 - Recr. 1990 or after					logOR ORR	logHR OS	-0.01 NR NR	0.28 0.28 0.24
	Lung (NSCLC)[58]						lnOR ORR	lnHR OS	-0.02	-0.13
	Various (immuno)[67]						logOR ORR	logHR OS	-0.13	-0.26
	Colorectal[31]	- All - Anti-angio - Non-anti-angio					rr ORR	HR OS	NR	-0.03 -0.11 -0.06
	Renal cell[32]						ln rr ORR	-lnHR OS	-0.11	0.30
	Biliary tract[56]	- Chemo - Gemcitabine - Targeted					Ratio of ORR	log ratio of median OS	0.01 0.02 0.12	0.28 0.27 0.16

Surrogate relationship	Cancer types and refs	Subgroup	Based on difference in response				Based on relative risk or odds ratio for response			
			Surrogate	Final	Intercept	Slope	Surrogate	Final	Intercept	Slope
	Lung (SCLC)[41]	- All - Pub. 1990-1996 - Pub. 1997-2008					rr ORR	Diff median OS	0.00 0.00 0.00	0.06 0.04 0.09
CR to PFS	NHL[69]						logOR CR 30mo	logHR PFS	-0.09	-0.64
	NHL[69]						logOR CR 24mo	logHR PFS	0.04	-0.73
CR to OS	Breast[36]	- All - Recr. pre-1990 - Recr. 1990 or after					logOR CR	logHR OS	-0.01 NR NR	0.13 0.09 0.16

Anti-angio, anti-angiogenic; CR, complete response; diff, difference; HR, hazard ratio; ln, natural logarithm; log, logarithm; NHL, non-Hodgkin's lymphoma; NR, not reported; NSCLC, non-small cell lung cancer; OR, odds ratio; ORR, overall response rate; OS, overall survival; PFS, progression-free survival; pub, published; recr, recruited; rr, relative risk; SCLC, small cell lung cancer.

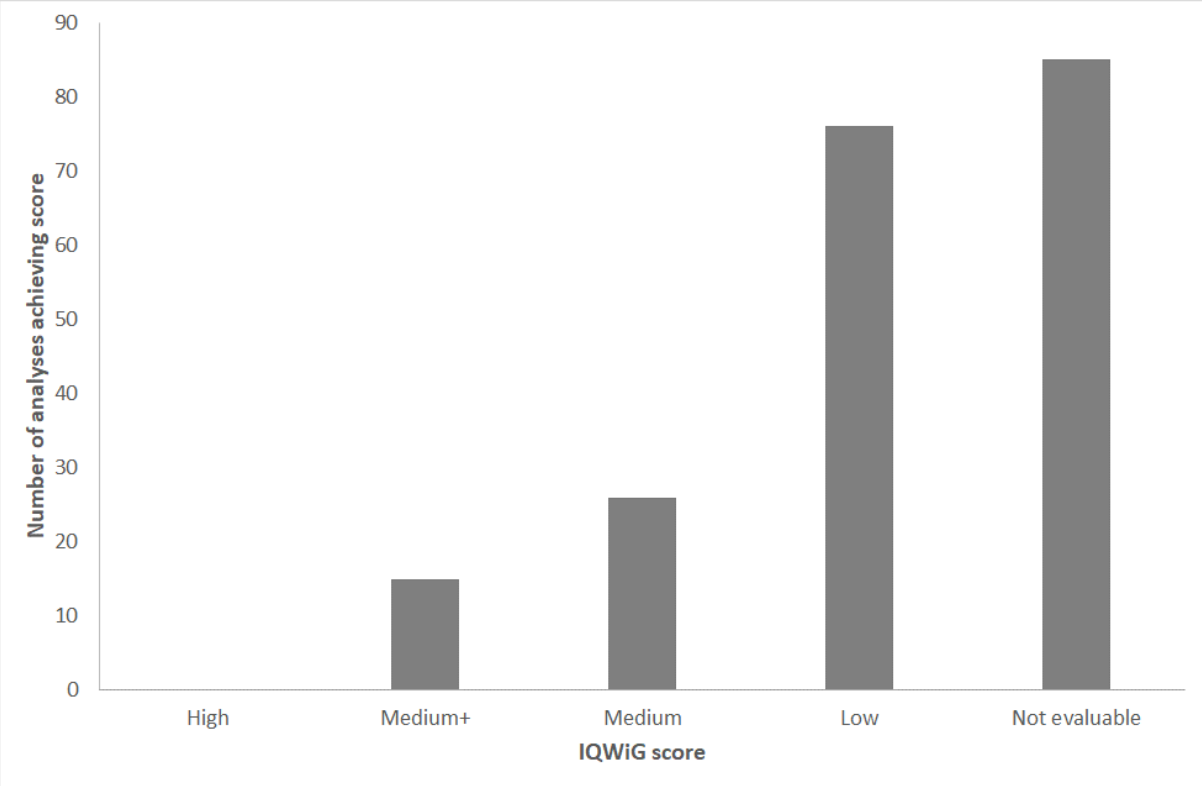
19

Supplementary Table 12: Surrogate threshold effect (STE) data reported per study

Surrogate relationship	Cancer types and refs	Based on difference in response			Based on odds ratio for response		
		Surrogate	**Final**	**STE**	**Surrogate**	**Final**	**STE**
ORR to PFS	Various[77]	Diff ORR	HR PFS	15%			
ORR to OS	Colorectal[26]				OR ORR	OR OS	0.28
	NSCLC[39]	Diff ORR Diff ORR	HR OS Diff median OS	55% 41%			
	Various[77]	Diff ORR	HR OS	21%			
CR to PFS	NHL[69]				OR CR 30mo	HR PFS	1.56

CR, complete response; diff, difference; HR, hazard ratio; NHL, non-Hodgkin's lymphoma; NSCLC, non-small cell lung cancer; OR, odds ratio; ORR, overall response rate; OS, overall survival; PFS, progression-free survival; STE, surrogate threshold effect.

Supplementary Figure 1: IQWiG scores for strength of association across all 202 analyses (within 63 included studies)



Supplementary Figure 2: BSES2 scores for strength of association across all 202 analyses (within 63 included studies)