Appendices

Appendix A – Search terms

Database	Search terms
Medline	1. Ipilimumab; 2. MDX-010; 3. MDX-101; 4. Yervoy; 5. BMS-734016; 6. Nivolumab; 7. ONO-4538; 8. BMS-936558; 9. MDX-1106; 10. Opdivo; 11. Pembrolizumab; 12. MK-4375; 13. Lambrolizumab; 14. Keytruda; 15. Checkpoint inhib*; *16. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15;* 17. *Melanoma or Melanoma/ or Melanoma skin cancer*; 18. Malignant melanoma; 19. *Skin tumor or Skin cancer*; 20. *Skin neoplasm or Skin neoplasm/;* 21. Skin carcinoma; 22. *17 or 18 or 19 or 20 or 21*; 23. *16 and 22*
Embase	1. Ipilimumab or Ipilimumab/; 2. MDX-010; 3. MDX-101;4. Yervoy; 5. BMS-734016; 6. Nivolumab or Nivolumab/; 7. ONO-4538; 8. BMS-936558; 9. MDX-1106; 10. Opdivo; 11. Pembrolizumab or Pembrolizumab/; 12. MK-4375; 13. Lambrolizumab; 14. Keytruda; 15. Checkpoint inhib*; 16. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15; 17. Melanoma or Melanoma/; 18. Melanoma skin cancer or Melanoma skin cancer/; 19. Malignant melanoma; 20. Skin tumor or Skin tumor/; 21. Skin cancer or Skin cancer/; 22. Skin carcinoma or Skin carcinoma/; 23. Skin neoplasm*; 24. 17 or 18 or 19 or 20 or 21 or 22 or 23; 25. 16 and 24
Cochrane	1. Ipilimumab or MDX-010 or MDX-101 or Yervoy or BMS 734016; 2. Nivolumab or ONO-4538 or BMS936558 or MDX-1106 or Opdivo; 3. Pembrolizumab or MK-4375 or Lambrolizumab or Keytruda; 4. Checkpoint inhib*; 5. 1 or 2 or 3 or 4; 6. Melanoma/; 7. Melanoma or Melanoma skin cancer; 8. Malignant melanoma; 9. Skin neoplasm/; 10. Skin cancer or Skin tumor or Skin carcinoma or Skin neoplasm; 11. 6 or 7 or 8 or 9 or 10; 12. 5 and 11
Web of Science	1. Ipilimumab; 2. MDX-010; 3. MDX-101; 4. Yervoy; 5. BMS-734016; 6. Nivolumab; 7. ONO-4538; 8. BMS-936558; 9. MDX-1106; 10. Opdivo; 11. Pembrolizumab; 12. MK-4375; 13. Lambrolizumab; 14. Keytruda; 15. Checkpoint inhib*; 16. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15; 17. Melanoma or Melanoma skin cancer or Malignant melanoma or Skin tumor or Skin cancer or Skin neoplasm* or Skin carcinoma; 18. 16 and 17

Table 1. Search strategy for the databases

The search terms used on the four databases, with each number being an individual search and an italicized region representing a group of similar search terms referring to the same drug or disease. ‘OR’ & ‘AND’ were used to connect separate searches.

/ MeSH search term.

* Open-ended search.

Appendix B – Participant data

Author	Mean age	Female – (%)	ECOG Performance status † – no. (%)	Metastasis stage – no. (%)	Lactate dehydrogenase levels – no. (%)	Previous systemic therapy
Hodi	56.2	40.7	0 – 374 (55.3) 1 – 291 (43.0) 2 – 9 (1.3) 3 – 1 (0.1) Unknown – 1 (0.1)	M0 – 10 (1.5) M1a – 62 (9.2) M1b – 121 (17.9) M1c – 483 (71.4)	≤ULN – 417 (61.7%) >ULN – 254 (37.6) Unknown – 5 (0.7)	Yes (chemotherapy or IL-2)
Larkin	60.0	35.4	0 – 692 (73.2) 1 – 251 (26.6) 2 – 1 (0.1) Not reported – 1 (0.1)	M0, M1a, M1b – 397 (42.0) M1c – 548 (58.0)	≤ULN – 589 (62.3) >ULN – 341 (36.1) Unknown – 15 (1.6)	No
Postow	65 *	33.1	0 – 116 (81.7) 1 – 24 (16.9) ≥2 – 2 (1.4)	M0 – 13 (9.2) M1a – 23 (16.2) M1b – 39 (27.5) M1c – 65 (45.8) Not reported – 2 (1.4)	≤ULN – 106 (74.6) >ULN – 35 (24.6) Unreported – 1 (0.7)	No

Author	Mean age	Female – (%)	ECOG Performance status † – no. (%)	Metastasis stage – no. (%)	Lactate dehydrogenase levels – no. (%)	Previously systemic therapy
Ribas	61.7 *	39.4	0 – 295 (54.6) 1 – 243 (45.0) Not reported – 2 (0.4)	M0 – 4 (0.7) M1a – 37 (6.9) M1b – 54 (10.0) M1c – 445 (82.4)	Normal – 311 (57.6) ≥110% ULN) – 218 (40.4) Unknown – 11 (2.0)	Yes (ipilimumab ± BRAF and/or MEK inhibitor and/or chemotherapy)
Robert (2011)	56.9	40.0	0 – 356 (70.9) 1 – 146 (29.1)	M0 – 14 (2.8) M1a – 80 (15.9) M1b – 126 (25.1) M1c – 282 (56.2)	≤ULN – 297 (59.2%) >ULN – 203 (40.4) Unknown – 2 (0.4)	No ** (Adjuvant therapy – 133 (26.5))
Robert (2015)	65.0 ‡	41.1	0 – 269 (64.4) 1 – 144 (34.4) 2 – 4 (1.0)	M0, M1a, M1b – 163 (39.0) M1c – 255 (61.0)	≤ULN – 245 (58.6%) >ULN - 153 (36.6) Not reported – 20 (4.8)	No ** (Adjuvant – 68 (16.3)) Neoadjuvant – 2 (0.5))
Weber	60.0 *	35.6	0 – 246 (60.7) 1 – 158 (39.0) Not reported – 1 (0.2)	M1c – 305 (75.3) Not reported – 100 (24.7)	>ULN – 185 (45.7) Not reported – 220 (54.3)	Yes (ipilimumab, ± BRAF inhibitor and/or chemotherapy)

Table 2. Baseline participant demographics and disease status

An overview of baseline participant data on demographics (mean age of study population, and proportion of female participants), and disease status (ECOG performance status, metastasis stage, and levels of lactate dehydrogenase). Previous treatments in the study populations are also shown in the last column.

* The median rather than the mean age is reported.

† Eastern Cooperative Oncology Group (ECOG) performance-status scores ranges from 0 to 5, where 0 is no symptoms, 1 is symptomatic but completely ambulatory, and 2 and 3 is symptomatic and in bed during the day <50% and >50%, respectively.

‡ The mean of each arm was manually calculated as no data was given for the whole study population.

** Patients were previously untreated, but for a group of patients that had received past adjuvant or neoadjuvant therapy.

Online supplementary material

Section A – Study quality assessment

The quality of all included studies was assessed using the 2010 CONSORT checklist. For each item, the studies were scored as 'done', 'not done', or 'not applicable', as seen in the example for the Hodi study[15], in Table 5 below. From this an overall score of the study quality was calculated based on the equation:

$$\frac{items\ done}{(total\ items - not\ applicable)} \times 100$$

The overall mean, minimum and maximum scores were calculated, and items consistently done poorly were noted. The results were used to test for the strength of correlation between study quality and primary efficacy outcome in order to assess whether poor quality studies may have biased the results of the meta-analysis, as described in the Results section. Additionally, the effect that removing the poorest quality studies had on the heterogeneity measure (I^2) was determined.

Section/Topic	Item No	Checklist item	Reported on page No.
Title and abstract			
	1a	Identification as a randomized trial in the title	**X**
	1b	Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts) **X**	✓ 711
Introduction			
Background and objectives	2a	Scientific background and explanation of rationale	✓ 712
	2b	Specific objectives or hypotheses	**X**
Methods			
Trial design	3a	Description of trial design (such as parallel, factorial) including allocation ratio	✓ 713
	3b	Important changes to methods after trial commencement (such as eligibility criteria), with reasons	N/A
Participants	4a	Eligibility criteria for participants	✓ 712
	4b	Settings and locations where the data were collected	✓ 712

Interventions	5	The interventions for each group with sufficient details to allow replication, including how and when they were actually administered	✓ 713
Outcomes	6a	Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed	X
	6b	Any changes to trial outcomes after the trial commenced, with reasons	✓ 713
Sample size	7a	How sample size was determined	✓ 714
	7b	When applicable, explanation of any interim analyses and stopping guidelines	✓ 714
Randomization:			
Sequence generation	8a	Method used to generate the random allocation sequence	X
	8b	Type of randomization; details of any restriction (such as blocking and block size)	X
Allocation concealment mechanism	9	Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned	X
Implementation	10	Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions	X
Blinding	11a	If done, who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how	X
	11b	If relevant, description of the similarity of interventions	X
Statistical methods	12a	Statistical methods used to compare groups for primary and secondary outcomes	✓ 714
	12b	Methods for additional analyses, such as subgroup analyses and adjusted analyses	✓ 717

Results			
Participant flow (a diagram is strongly recommended)	13a	For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analyzed for the primary outcome	✓ 714
	13b	For each group, losses and exclusions after randomization, together with reasons	**X**
Recruitment	14a	Dates defining the periods of recruitment and follow-up	**X**
	14b	Why the trial ended or was stopped	N/A
Baseline data	15	A table showing baseline demographic and clinical characteristics for each group	✓ 715
Numbers analyzed	16	For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups	✓ 714
Outcomes and estimation	17a	For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval)	✓ 718
	17b	For binary outcomes, presentation of both absolute and relative effect sizes is recommended	✓ 718
Ancillary analyses	18	Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory	✓ 717
Harms	19	All important harms or unintended effects in each group (for specific guidance see CONSORT for harms)	✓ 720

Discussion			
Limitations	20	Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses	**X**
Generalizability	21	Generalizability (external validity, applicability) of the trial findings	✓ 712
Interpretation	22	Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence	✓ 721
Other information			
Registration	23	Registration number and name of trial registry	✓ 711
Protocol	24	Where the full trial protocol can be accessed, if available	✓ 712
Funding	25	Sources of funding and other support (such as supply of drugs), role of funders	✓ 721

Table 1. CONSORT checklist for Hodi (2010)

A sample 2010 CONSORT checklist as it appears for the Hodi study.[15] The 25 items are listed in the leftmost column, numbered, sub-categorized (37 items in total), and described in the subsequent columns, with the final column on the right containing a green tick and the page number if the criteria was met, a red cross if not met, or marked as N/A if not applicable.

Author	Fulfilled	Not fulfilled	Not Applicable	Score	Percentage
Hodi	23	12	2	23/35	65.7%
Larkin	22	13	2	22/35	62.9%
Postow	19	15	3	19/34	55.9%
Ribas	27	8	2	27/35	77.1%
Robert (2011)	21	14	2	21/35	60.0%
Robert (2015)	21	14	2	21/35	60.0%
Weber	26	8	3	26/34	76.5%
				Min.	55.9%
				Mean.	65.4%
				Max.	77.1%

Table 2. CONSORT study quality scores

The overall quality scores for each study as a fraction and a percentage is listed in the final two columns, along with the number of criteria that were fulfilled, not fulfilled, or not applicable. The minimum and maximum study scores, as well as the overall mean for all seven studies is listed in the bottom right.

Section B – Raw data

Author	Median overall survival – months (95% CI)	Median progression free survival – months (95% CI)	HR death (95% CI)	HR progression (95% CI)
Hodi	Ipi & gp100 – 10.0 (8.5 – 11.5) gp100 – 6.4 (5.5 – 8.7) Ipi – 10.1 (8.0 – 13.8)	Ipi & gp100 – 2.76 (2.73 – 2.79) gp100 – 2.76 (2.73 – 2.83) Ipi – 2.86 (2.76 – 3.02)	Ipi & gp100 vs. gp100 0.68 (0.55 – 0.85) Ipi vs. gp100 0.66 (0.51 - 0.87)	Ipi & gp100 vs. Ipi – 0.81 * Ipi vs. gp100 – 0.64 †
Larkin	*Data immature*	Nivo & Ipi – 11.5 (8.9 – 16.7) Nivo & placebo – 6.9 (4.3 – 9.5) Ipi & placebo – 2.9 (2.8 – 3.4)	*Data immature*	Nivo & Ipi vs. Ipi – 0.42 (0.31 – 0.57) Nivo vs. Ipi – 0.57 (0.43 – 0.76) Nivo & Ipi vs. Nivo – 0.74 (0.60 – 0.92)
Postow	*Not reported*	Ipi & Nivo – *Data immature* Ipi – 4.4 (2.8 – 5.7)	*Not reported*	0.40 (0.23 - 0.68)

Author	Median overall survival – months (95% CI)	Median progression free survival – months (95% CI)	HR death (95% CI)	HR progression (95% CI)
Ribas	*Data immature*	Pembro 2mg/kg – 5.4 (4.7 – 6.0 Pembro 10mg/kg – 5.8 (5.1 – 6.4) ICC – 3.6 (3.2 – 4.1) ‡§	*Data immature*	Pembro 2m/kg vs. chemotherapy – 0.57 (0.45 – 0.73) Pembro 10m/kg vs. chemotherapy – 0.50 (0.39 – 0.64) Pembro 10m/kg vs. Pembro 2mg/kg – 0.91 (0.71 – 1.16)
Robert	Ipi & Dacarb. – 11.2 (9.4 – 13.6) Dacarb. – 9.1 (7.8 – 10.5)	*Not reported*	0.72 (0.59 - 0.87)	0.76 (0.63 – 0.93)
Robert	Nivo – *Data immature* Dacarb. – 10.8 (9.3 – 12.1)	Nivo – 5.1 (3.5 – 10.8) Dacarb. – 2.2 (2.1 – 2.4)	0.42 (0.25 - 0.73) §	0.43 (0.34 – 0.56)
Weber	*Data immature*	Nivo – 4.7 (2.3 – 6.5) ICC – 4.2 (2.1 – 6.3) ‡§	*Data immature*	0.82 (0.32 – 2.05) **

Table 3. Primary outcome raw data on survival

The raw data as reported in the included studies for median overall, and median progression-free survival in months, and hazard ratios for death, and progression for each study and treatment arm. Ipo, nivo, pembro, and dacarb are short for ipilimumab, nivolumab, pembrolizumab, and dacarbazine, respectively.

* p <0.05

† p <0.001

‡ Data from only 182 / 405 patients was reported

§ 99.79% confidence intervals

** 99.99% confidence intervals

Author	BORR – No. objective responses / Total no. patients (%)
Hodi	Ipilimumab & gp100 – 23/403 (5.7%) gp100 – 2/136 (1.5%) Ipilimumab – 15/137 (10.9%)
Larkin	Combination – 181/314 (57.6%) Nivolumab – 138/316 (43.7%) Ipilimumab – 60/315 (19.0%)
Postow *	Ipilimumab & Nivolumab – 56/95 (58.9%) Ipilimumab – 5/47 (10.6%)
Ribas	Pembrolizumab 2mg/kg – 38/180 (21.1%) Pembrolizumab 10mg/kg – 46/181 (25.4%) ICC – 8/179 (4.5%)
Robert	Ipilimumab & Dacarbazine – 38/250 (15.2%) Dacarbazine – 26/252 (10.3%)
Robert	Nivolumab – 84/210 (40.0%) Dacarbazine – 29/208 (13.9%)
Weber	Nivolumab – 38/122 (31.1%) ICC – 5/60 (8.3%)

Table 4. Secondary outcome raw data on tumor response

The raw data as reported in the included studies for best overall response rate (BORR), ie the number of patients in each study and treatment arm that achieve an objective response (complete or partial response) as a fraction of the total number of patients.

* Combined data for BRAF wild-type tumors, and BRAF V600 mutation-positive tumors, which was reported separately in the study.

Author	Discontinuations due to adverse events – No. events / Total (%)	Discontinuations due to treatment-related adverse events – No. events / Total (%)
Hodi *	*Not reported*	*Not reported*
Larkin	*Not reported*	Combination – 114/313 (36.4%) Nivolumab – 24/313 (7.7%) Ipilimumab – 46/311 (14.8%)
Postow	*Not reported*	Ipilimumab & Nivolumab – 44/94 (46.8%) Ipilimumab – 8/46 (17.4%)
Ribas	Pembrolizumab 2mg/kg – 21/178 (11.8%) Pembrolizumab 10mg/kg – 24/179 (13.4%) ICC – 18/171 (10.5%)	Pembrolizumab 2mg/kg – 4/178 (2.2%) Pembrolizumab 10mg/kg – 12/179 (6.7%) ICC – 10/171 (5.8%)
Robert	*Not reported*	Ipilimumab & Dacarbazine – 89/247 (36.0%) Dacarbazine – 10/251 (4.0%)
Robert	Nivolumab – 14/206 (6.8%) Dacarbazine – 24/205 (11.7%)	*Not reported*
Weber	*Not reported*	Nivolumab – 7/268 (2.6%) ICC – 7/102 (6.9%)

Table 5. Secondary outcome raw data on tolerability

The raw data as reported in the included studies for the secondary outcome on tolerability, with rates of discontinuation due to adverse events, and treatment-related adverse events for each study and treatment arm listed.
* Study reported neither tolerability endpoint and is thus not included in the meta-analysis.

Section C – Risk of bias assessment

Bias	Author's judgment	Support for judgment
Random sequence generation (selection bias)	Low risk	"Patients were randomly assigned to one of three study groups" Comment: Probably done.
Allocation concealment (selection bias)	Unclear risk	"The Biostatistics group in Medarex will provide a centralized randomization list to Clinical Operations using SAS procedure PROC PLAN. The randomization will be performed in two separate stages using different block sizes for different treatment allocation ratios." Comment: Unclear who performed randomization and what the method was used.
Blinding of participants and personnel (performance bias)	Low risk	"Placebo will be utilized for both MDX-010 and melanoma peptide vaccine. The melanoma peptide vaccine (placebo and active) will be delivered via masked syringe by s.c. injection." "All Study Site personnel, patients, and Medarex, Inc. personnel involved in the study...will remain blinded to treatment assignment during the course of the study." Comment: Probably done.

Blinding of outcome assessment (detection bias)	Low risk	"Tumor responses were determined by the investigators with the use of modified WHO criteria to evaluate bidimensionally measurable lesions." "The IRC will be blinded to patient dosage group assignments…The IRC will be comprised of at least 2 radiologists or oncologists experienced in tumor imaging and assessment." Comment: IRC assessed tumor-response data. Other relevant personnel were also blinded.
Incomplete outcome data (attrition bias)	Unclear risk	"Efficacy analyses were performed on the intention-to-treat population, which included all patients who had undergone randomization (676 patients). The safety population included all patients who had undergone randomization and who had received any amount of study drug (643 patients)" "Of the 143 patients who could not be evaluated for a response, 33 patients did not receive any study drug and 110 patients did not have baseline or week-12 tumor assessments (or both)" Comment: 143 patients were not evaluated for BORR, of which 110 patients, without further explanation, were said to lack data to compare with.
Selective reporting (reporting bias)	Unclear risk	Comment: Changed the primary outcome from BORR to overall survival in January 2009. Reported all specified outcomes.
Other bias	Unclear risk	"Funded by Medarex and Bristol-Myers Squibb" "The trial was designed jointly by the senior academic authors and the sponsors, Medarex and Bristol-Myers Squibb. Data were collected by the sponsors and analyzed in collaboration with the senior academic authors." Comment: Lead authors received consulting fees, grants, honoraria, and fees from BMS (patent holders for Ipilimumab).

Table 6. Risk of bias data for Hodi (2010)

The primary risk of bias assessment for the Hodi study[15], listing for each of the seven study domains firstly, the review author's judgment of the overall risk of bias (low, unclear, or high risk of bias), and secondly, the support for judgment consisting of extracts from the study or its supplementary material as well as comments made by the review author. An unclear risk of bias was defined as a risk of bias that was greater than low, but not sufficient to be considered high.

Bias	Author's judgment	Support for judgment
Random sequence generation (selection bias)	Low risk	"Enrolled patients were randomly assigned" Comment: Probably done.
Allocation concealment (selection bias)	Low risk	"The randomization procedures will be carried out via permuted blocks within each stratum. The exact procedures for using the IVRS will be detailed in the IVRS manual." Comment: Probably done.
Blinding of participants and personnel (performance bias)	Low risk	"For subjects who are receiving treatment and have not progressed, the Sponsor, subjects, investigator and site staff will be blinded to the study drug administered" "Each investigative site must assign an unblinded pharmacist/designee, and an unblinded site monitor will be assigned by sponsor to provide oversight of drug supply and other unblinded study documentation. "The Sponsor's central protocol team (including but not limited to clinical, statistics, data management) will remain blinded." Comment: Used placebo and blinded staff.

Blinding of outcome assessment (detection bias)	Unclear risk	"Tumor assessments for ongoing study treatment decisions will be completed by the investigator using RECIST (Response Evaluation Criteria in Solid Tumors) 1.1criteria. Radiographic images will be collected for independent radiological review committee tumor assessment." Comment: Unclear whether the investigator remained blinded. No description of the independent radiological review committee.
Incomplete outcome data (attrition bias)	Unclear risk	Comment: BOR could not be determined in 78 patients, with no explanation as to why.
Selective reporting (reporting bias)	Low risk	Comment: Reported specified outcomes with the exception of median overall survival (not mature) and PD-L1 expression as a predictive marker of efficacy.
Other bias	Unclear risk	"Funded by Bristol-Myers Squibb" "The trial was designed as a collaboration between the senior academic authors and the sponsor, Bristol-Myers Squibb. Data were collected by the sponsor and analyzed in collaboration with all the authors." Comment: BMS holds the patent for Ipilimumab. Authors declared receiving funds, grants, and honoraria from pharmaceutical industry, including BMS.

Table 7. Risk of bias data for Larkin (2015)

The primary risk of bias assessment for the Larkin study[35], listing for each of the seven study domains firstly, the review author's judgment of the overall risk of bias (low, unclear, or high risk of bias), and secondly, the support for judgment consisting of extracts from the study or its supplementary material as well as comments made by the review author. An unclear risk of bias was defined as a risk of bias that was greater than low, but not sufficient to be considered high.

Bias	Authors' judgment	Support for judgment
Random sequence generation (selection bias)	Low risk	"We randomly assigned patients in a 2:1 ratio" Comment: Probably done.
Allocation concealment (selection bias)	Low risk	"Enrolled subjects that have met all eligibility criteria will be ready to be randomized through the IVRS" "The randomization procedures will be carried out via permuted blocks within each stratum." Comment: Probably done.
Blinding of participants and personnel (performance bias)	Low risk	"The Sponsor, subjects, investigator and site staff will be blinded to the study drug administered" "Each investigative site must assign an unblinded pharmacist/designee, and an unblinded site monitor will be assigned by sponsor to provide oversight of drug supply and other unblinded study documentation." "In the ipilimumab-monotherapy group, the same dosing schedule was used, except that nivolumab was replaced with matched placebo" Comment: Used placebo and blinded staff.

Blinding of outcome assessment (detection bias)	Unclear risk	"The best overall response was assessed by the investigator with the use of the Response Evaluation Criteria in Solid Tumors" "An independent radiology review committee was established to provide a sensitivity assessment of objective responses," Comment: Unclear whether the investigator remained blinded. No description of the independent radiology review committee.
Incomplete outcome data (attrition bias)	Unclear risk	Comment: BOR could not be determined in 18, with no explanation as to why. 1 patient's LDH, and 1 patient's history of brain metastasis was not recorded.
Selective reporting (reporting bias)	Low risk	Comment: All endpoints reported.
Other bias	Unclear risk	"Data were collected by the sponsor, Bristol-Myers Squibb, and were analyzed in collaboration with the authors." Comment: Study funded by BMS (patent holders of Ipilimumab). Authors declared receiving funds, grants, and honoraria from pharmaceutical industry, including BMS (patent-holders).

Table 8. Risk of bias data for Postow (2015)

The primary risk of bias assessment for the Postow study[36], listing for each of the seven study domains firstly, the review author's judgment of the overall risk of bias (low, unclear, or high risk of bias), and secondly, the support for judgment consisting of extracts from the study or its supplementary material as well as comments made by the review author. An unclear risk of bias was defined as a risk of bias that was greater than low, but not sufficient to be considered high.

Bias	Author's judgment	Support for judgment
Random sequence generation (selection bias)	Low risk	"We randomly assigned (1:1:1) patients in a block size of six" Comment: Probably done.
Allocation concealment (selection bias)	Low risk	"Block randomization with a block size of six in each stratum was used. After all screening procedures were complete, a centralized interactive voice-response system with or without web functionality was used to allocate patients to treatment." Comment: Probably done.
Blinding of participants and personnel (performance bias)	High risk	"Individual treatment assignment between pembrolizumab and chemotherapy was open label; investigators and patients were masked to assignment to pembrolizumab dose. A designated pharmacist at each site who was unmasked prepared the pembrolizumab dose so that it could be administered to the patient in a masked fashion." "The sponsor was masked to all treatment assignments in the statistical analyses, as well as treatment-level analysis results." Comment: No placebo used, and assignment to chemotherapy or pembrolizumab was open-label.

Blinding of outcome assessment (detection bias)	Unclear risk	"All scans were evaluated by independent central review. The independent radiologists were masked to treatment assignments, identifying patient characteristics, and investigator-assessed findings." Comment: Outcomes also assessed by investigator for "sensitivity analysis", but these results were reported separately. No description of independent central review committee.
Incomplete outcome data (attrition bias)	Unclear risk	Comment: BOR was not evaluable in 71 patients, a fraction of which was due to patients being "withdrawn by investigator" with no further explanation. 16 patients discontinued their assigned treatment due to "physician decision" with no further explanation.
Selective reporting (reporting bias)	Low risk	Comment: All outcomes reported, with the exception of median overall survival (not mature), and time from BOR to disease progression.
Other bias	Unclear risk	"Merck Sharp & Dohme, a subsidiary of Merck & Co, sponsored this study". Comment: Pharmaceutical company also helped design study and collect data. Authors declared receiving funds, grants, and honoraria from pharmaceutical industry, including Merck & Co who holds the pembrolizumab patent.

Table 9. Risk of bias data for Ribas (2015)

The primary risk of bias assessment for the Ribas study[34], listing for each of the seven study domains firstly, the review author's judgment of the overall risk of bias (low, unclear, or high risk of bias), and secondly, the support for judgment consisting of extracts from the study or its supplementary material as well as comments made by the review author. An unclear risk of bias was defined as a risk of bias that was greater than low, but not sufficient to be considered high.

Bias	Author's judgment	Support for judgment
Random sequence generation (selection bias)	Low risk	"We randomly assigned 502 patients" Comment: Probably done.
Allocation concealment (selection bias)	Low risk	"To randomize an eligible patient, the unblinded pharmacist will call IVRS to obtain a treatment assignment." Comment: Used an interactive voice response system. No further information on design of stratum or blocks.
Blinding of participants and personnel (performance bias)	Low risk	"The Sponsor, CRO, patients, Investigator and site staff will be blinded to the ipilimumab dose (i.e., placebo or 10 mg/kg). The local pharmacists in addition to a pharmacy - based CRO monitor will be unblinded. The DMC will also be unblinded to permit a real-time ongoing assessment of safety and efficacy." Comment: Placebo used, and all relevant personnel were blinded, no description of the presentation of the placebo.

Blinding of outcome assessment (detection bias)	Unclear risk	"Tumor assessments were performed by the local investigator and by a central independent review committee." "All efficacy end points (except survival) were based on assessments performed by the independent review committee, whose members were not aware of the treatment assignments." "For the purpose of final analysis of study results, an IRC will review all images from all time points for all patients and assess response parameters as specified." Comment: Probably done, no description of who made up the IRC.
Incomplete outcome data (attrition bias)	Unclear risk	Comment: 101/502 patients had their "response not evaluated" for BOR, due to some lacking a baseline and/or follow-up scan. Two patients had unknown LDH levels.
Selective reporting (reporting bias)	Low risk	Comment: Changed primary end point from progression-free survival to overall survival. Reported all specified outcomes with the exception of time to a response.
Other bias	Unclear risk	"Funded by Bristol-Myers Squibb" "The trial was designed jointly by the senior academic authors and the sponsor, Bristol-Myers Squibb. Data were collected by the sponsor and analyzed in collaboration with the senior academic authors" Comment: BMS hold the patent for Ipilimumab. Authors declared receiving funds, grants, and honoraria from pharmaceutical industry, including BMS.

Table 10. Risk of bias data for Robert (2011)

The primary risk of bias assessment for the Robert (2011) study[21], listing for each of the seven study domains firstly, the review author's judgment of the overall risk of bias (low, unclear, or high risk of bias), and secondly, the support for judgment consisting of extracts from the study or its supplementary material as well as comments made by the review author. An unclear risk of bias was defined as a risk of bias that was greater than low, but not sufficient to be considered high.

Bias	Author's judgment	Support for judgment
Random sequence generation (selection bias)	Low risk	"We randomly assigned 418 previously untreated patients" Comment: Probably done.
Allocation concealment (selection bias)	Low risk	"The subject number will be assigned through an interactive voice response system (IVRS)" "The randomization procedures will be carried out via permuted blocks within each stratum." Comment: Probably done.
Blinding of participants and personnel (performance bias)	Low risk	"The Sponsor, subjects, investigator and site staff will be blinded to the study drug administered" "Each investigative site must assign an unblinded pharmacist/designee, and an unblinded site monitor will be assigned to provide oversight of drug supply and other unblinded study documentation." Comment: Used placebo, and blinded relevant personnel, but no description of the presentation of the placebo presentation.

Blinding of outcome assessment (detection bias)	High risk	"The best overall response was assessed by the investigator with the use of the Response Evaluation Criteria in Solid Tumors" "The duration of investigator-assessed progression-free survival (PFS)" Comment: No mention of an independent review committee, nor whether the investigators remained masked during outcome assessment.
Incomplete outcome data (attrition bias)	Unclear risk	Comment: BOR could not be determined in 54/418 patients, without further explanation as to why. LDH and BRAF status not reported 20 and 12 patients, respectively.
Selective reporting (reporting bias)	Low risk	Comment: Reported all specified outcomes with the exception of median overall survival (not mature) and PD-L1 expression as a predictive marker of efficacy.
Other bias	Unclear risk	"Funded by Bristol-Myers Squibb" "Data were collected by the sponsor, Bristol-Myers Squibb, and analyzed in collaboration with the academic authors." Comment: BMS hold the patent for Nivolumab. Authors declared receiving funds, grants, and honoraria from pharmaceutical industry, including BMS.

Table 11. Risk of bias data for Robert (2015)

The primary risk of bias assessment for the Robert (2015) study[29], listing for each of the seven study domains firstly, the review author's judgment of the overall risk of bias (low, unclear, or high risk of bias), and secondly, the support for judgment consisting of extracts from the study or its supplementary material as well as comments made by the review author. An unclear risk of bias was defined as a risk of bias that was greater than low, but not sufficient to be considered high.

Bias	Author's judgment	Support for judgment
Random sequence generation (selection bias)	Low risk	"Participating investigators randomly assigned (with an interactive voice response system) patients" Comment: Probably done.
Allocation concealment (selection bias)	Low risk	"We used permuted blocks (block size of six) within each stratum." Comment: Used an IVRS.
Blinding of participants and personnel (performance bias)	High risk	"Treatment was given open-label because of the choices available to the investigators in the ICC group" Comment: An open-label study.

Blinding of outcome assessment (detection bias)	Unclear risk	"Tumor assessments were done centrally by radiologists on an independent review committee who were masked to patients' treatment assignments." "Confirmed response by independent radiology review committee per Response Evaluation Criteria in Solid Tumors" Comment: No description of the IRC.
Incomplete outcome data (attrition bias)	Unclear risk	Comment: Reports data from only 182 / 405 patients – number of patients "who had been randomized at the point of the first planned assessment of objective responses". In 23/167 they were "unable to establish" BOR, due to lack of scan at 9 months without any further explanation. Comment: Unclear whether remaining data will be published.
Selective reporting (reporting bias)	Unclear risk	Comment: Reported some specified outcomes but not all, specifically: PD-L1 expression as a predictive biomarker for objective response, overall survival (not mature), and health-related Quality of life.
Other bias	Unclear risk	"Funding Bristol-Myers Squibb" "Data collected by the funder were analyzed in collaboration with all authors." Comment: BMS hold the patent for Nivolumab. Authors declared receiving funds, grants, and honoraria from pharmaceutical industry, including BMS.

Table 12. Risk of bias data for Weber (2015)

The primary risk of bias assessment for the Weber study[22], listing for each of the seven study domains firstly, the review author's judgment of the overall risk of bias (low, unclear, or high risk of bias), and secondly, the support for judgment consisting of extracts from the study or its supplementary material as well as comments made by the review author. An unclear risk of bias was defined as a risk of bias that was greater than low, but not sufficient to be considered high.

Section D – Publication bias assessment

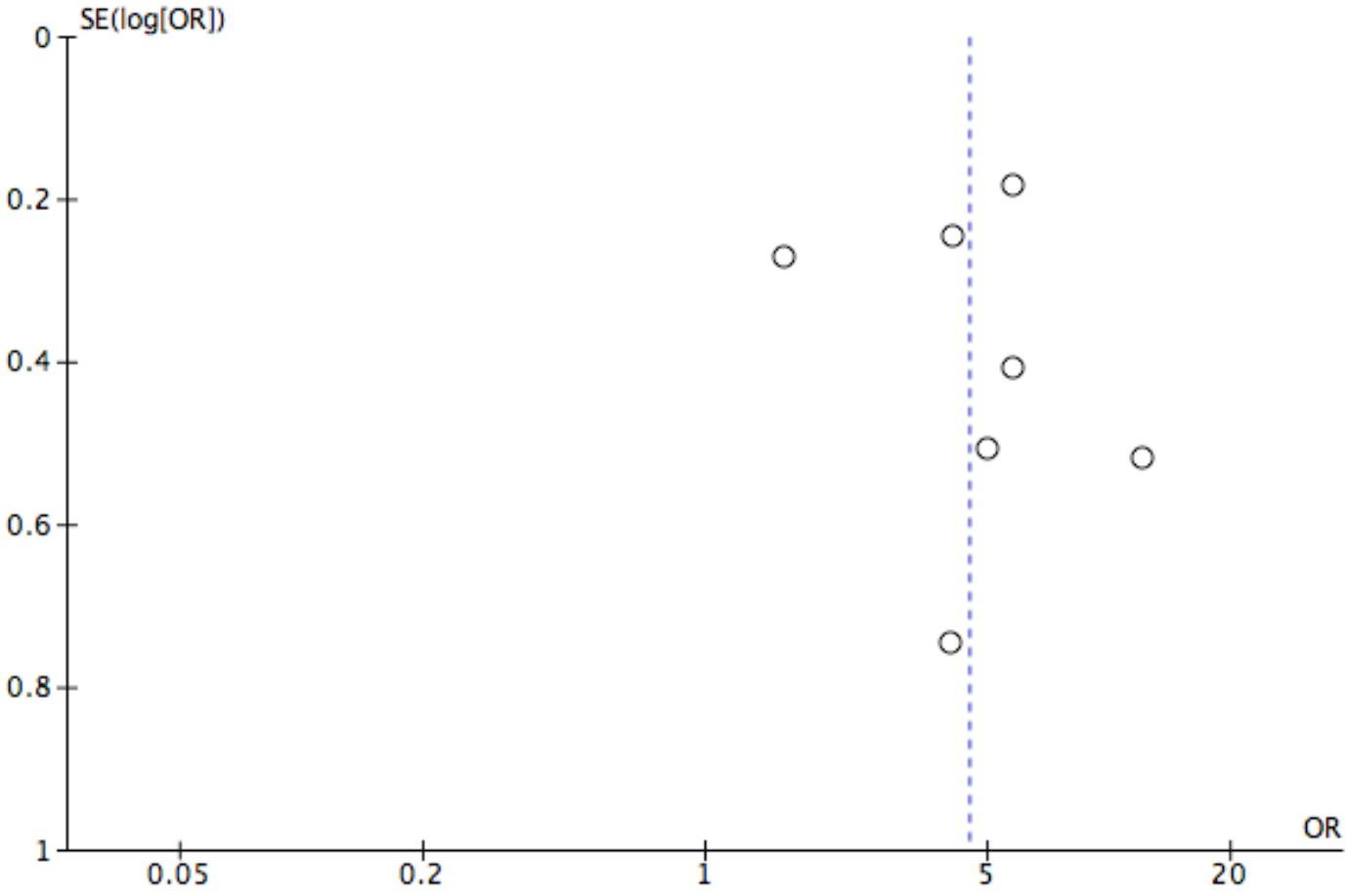

Figure 1. Funnel Plot for the secondary outcome analysis on tumor response

The funnel plot for the secondary outcome analysis on tumor response, showing each study as a black circle, with the odds ratio for best overall response rate (BORR) along the x-axis, and the standard error of the natural log of the odds ratio on the y-axis. The smaller the SE (log [Odds Ratio]), the more reliable the result from that studies is, meaning less reliable studies will be found closer to the x-axis. There is an even spread of studies on either side of the vertical blue line representing the overall effect estimate (OR = *4.48*).

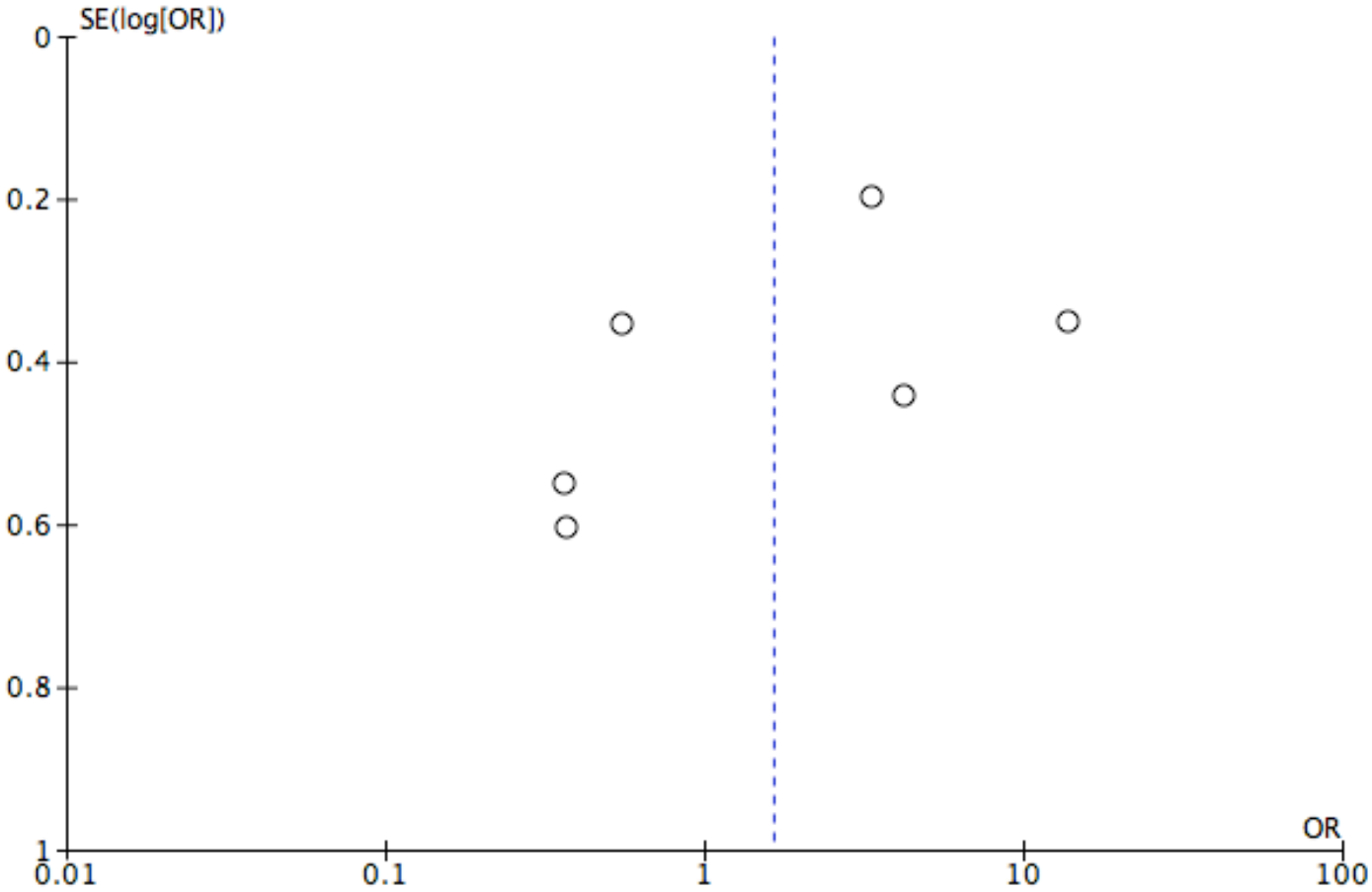

Figure 2. Funnel Plot for the secondary outcome analysis on tolerability

The funnel plot for the secondary outcome analysis on tolerability, showing each study as a black circle, with the odds ratio for rates of discontinuations due to adverse and treatment-related adverse events along the x-axis, and the standard error of the natural log of the odds ratio on the y-axis. The smaller the SE (log [Odds Ratio]), the more reliable the result from that studies is, meaning less reliable studies will be found closer to the x-axis. There is an even spread of studies on either side of the vertical blue line representing the overall effect estimate (OR = *1.63*).