First-line systemic therapy for advanced gastric cancer: a systematic review and network meta-analysis

Ji Cheng et al

Search strategies: Page 2-3
eTable 1. Studies eligible for exclusion criteria: Page 4
eTable 2. Risk of bias assessment of studies for general analysis (unselected population): Page 5-26
eTable 3. Comparisons between direct and indirect results regarding overall survival (general analysis): Page 27
eTable 4. Risk of bias assessment of studies for additional analysis (unselected population): Page 28-32
eTable 5. Risk of bias assessment of studies for patients with specific positivity: Page 33-34
eFigure 1. Funnel plot of overall survival (general analysis): Page 35
eFigure 2. Network forest plot for progression-free survival (general analysis): Page 36
eFigure 3. Network forest plot for objective response rate (general analysis): Page 37
eFigure 4. Network forest plot for hematological adverse events (general analysis): Page 38
eFigure 5. Network forest plot for non-hematological adverse events (general analysis): Page 39
eFigure 6. Network forest plot for overall survival (additional analysis): Page 40
eFigure 7. Network league table for overall survival (additional analysis): Page 41
eFigure 8. Network forest plot for progression-free survival (additional analysis): Page 42
eFigure 9. Network league table for progression-free survival (additional analysis): Page 43
eFigure 10. Network forest plot for objective response rate (additional analysis): Page 44
eFigure 11. Network league table for objective response rate (additional analysis): Page 45
eFigure 12. Network forest plot for hematological adverse events (additional analysis): Page 46
eFigure 13. Network league table for hematological adverse events (additional analysis): Page 47
eFigure 14. Network forest plot for non-hematological adverse events (additional analysis): Page 48
eFigure 15. Network league table for non-hematological adverse events (additional analysis): Page 49
Supplementary figure legends: Page 50-51
Supplementary reference: Page 52-53
1. PubMed
Search terms:
   #1 Search (((((gastric) OR stomach) OR esophagogastric) OR oesophagogastric) OR gastroesophageal) OR gastrooesophageal; Items found: 401807
   #2 Search (((cancer) OR carcinoma) OR adenocarcinoma; Items found: 3792614
   #3 Search ((((((advanced) OR metastatic) OR metastasis) OR recurrent) OR unresectable) OR inoperable) OR incurable) OR palliative; Items found: 1864980
   #4 Search (((randomized) OR randomised) OR randomly) OR random; Items found: 1213016
   #5 #1 AND #2 AND #3 AND #4; **Items found: 2659**
   (All fields; No limitations)

2. Web of Science (Core Collection)
Search terms:
   #1 TOPIC: (gastric) OR TOPIC: (stomach) OR TOPIC: (esophagogastric) OR TOPIC: (oesophagogastric) OR TOPIC: (gastroesophageal); Results: 351831
   #2 TOPIC: (cancer) OR TOPIC: (carcinoma) OR TOPIC: (adenocarcinoma); Results: 2594760
   #3 TOPIC: (advanced) OR TOPIC: (metastatic) OR TOPIC: (metastasis) OR TOPIC: (recurrent) OR TOPIC: (unresectable) OR TOPIC: (inoperable) OR TOPIC: (incurable) OR TOPIC: (palliative); Results: 1820659
   #4 TOPIC: (randomized) OR TOPIC: (randomised) OR TOPIC: (randomly) OR TOPIC: (random); Results: 1687422
   #5 #1 AND #2 AND #3 AND #4; **Results: 4074**
   (Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, BKCI-S, BKCI-SSH, ESCI, CCR-EXPANDED, IC Timespan=All years)

3. Cochrane Central Register of Controlled Trials
Search terms:
   #1 (gastric) OR (stomach) OR (esophagogastric) OR (oesophagogastric) OR (gastroesophageal OR gastrooesophageal); Results: 28934
   #2 (cancer) OR (carcinoma) OR (adenocarcinoma); Results: 142541
   #3 (advanced) OR (metastatic OR metastasis) OR (recurrent) OR (unresectable OR inoperable OR incurable) OR (palliative); Results: 124475
   #4 (randomized) OR (randomised) OR (randomly) OR (random); Results: 873090
   #5 #1 AND #2 AND #3 AND #4; **Results: 2519**
   (All Text; All Dates; Search word variations)

4. Embase
Search terms:
   #1 gastric OR stomach OR esophagogastric OR oesophagogastric OR gastroesophageal OR gastrooesophageal; Results: 633503
   #2 cancer OR carcinoma OR adenocarcinoma; Results: 4132339
#3 advanced OR metastatic OR metastasis OR recurrent OR unresectable OR inoperable OR incurable OR palliative; Results: 2012469
#4 randomized OR randomised OR randomly OR random; Results: 1507296
#5 #1 AND #2 AND #3 AND #4; Results: 4961
(All fields; No limitations)

5. ASCO Meeting Library
Search terms:
(gastric OR stomach OR esophagogastric OR oesophagogastric OR gastroesophageal OR gastrooesophageal) AND (cancer OR carcinoma OR adenocarcinoma) AND (advanced OR metastatic OR metastasis OR recurrent OR unresectable OR inoperable OR incurable OR palliative) AND (randomized OR randomised OR randomly OR random)
Results: 531
(No filter)

6. ESMO Meeting Library (Due to the restricted amount and manner of search terms that were allowed to be inserted at a time, we therefore separated the search terms)
Search terms:
#1 gastric AND randomized (Filter: Abstract; Meeting Report); Results: 191
#2 gastric AND randomly (Filter: Abstract; Meeting Report); Results: 56
#3 stomach AND randomized (Filter: Abstract; Meeting Report); Results: 45
#4 stomach AND randomly (Filter: Abstract; Meeting Report); Results: 17
#5 esophagogastric AND randomized (Filter: Abstract; Meeting Report); Results: 17
#6 esophagogastric AND randomly (Filter: Abstract; Meeting Report); Results: 4
#7 gastroesophageal AND randomized (Filter: Abstract; Meeting Report); Results: 38
#8 gastroesophageal AND randomly (Filter: Abstract; Meeting Report); Results: 10
Total results: 378
(Word variations including “randomised” “oesophagogastric” and gastrooesophageal” have been automatically searched)
| Study          | Reasons of exclusion            |
|----------------|---------------------------------|
| Kim 2018       | Could not incorporate into network calculation |
| Huang 2013     | Could not incorporate into network calculation |
| Huang 2012     | Could not incorporate into network calculation |
| Cascinu 2011   | Could not incorporate into network calculation |
| Lee 2008       | Could not incorporate into network calculation |
| Duffour 2006   | Could not incorporate into network calculation |
| Ross 2002      | Could not incorporate into network calculation |
| Kondo 2000     | Could not incorporate into network calculation |
| Wils 1991      | Could not incorporate into network calculation |
| Bang 2017      | Sequential therapy              |
| Li 2017        | Sequential therapy              |
| Kim 2011       | Sequential therapy              |
| Gubanski 2010  | Sequential therapy              |
| Bramhall 2002  | Sequential therapy              |
| Study          | Random sequence generation          | Allocation concealment       | Blinding of participants and personnel | Blinding of outcome assessment | Incomplete outcome data                  | Selective reporting                    | Other sources of bias                      |
|---------------|-------------------------------------|------------------------------|---------------------------------------|-------------------------------|------------------------------------------|----------------------------------------|-------------------------------------------|
| Yamada 2018   | Unclear: No specific description    | Low: Central dynamic allocation | High: Open-label design               | Unclear: No specific description | Unclear: No specific description         | Low: All expected endpoints had been reported. | Unclear: No specific description          |
| Muro 2018     | Unclear: No specific description    | Unclear: No specific description | Low: Double-blind placebo-controlled trial (participants and investigators) | Unclear: No specific description | Unclear: No specific description         | Unclear: No specific description         | Unclear: No specific description          |
| Lu 2018       | Low: Minimization method            | Unclear: No specific description | High: Open-label design               | Unclear: No specific description | Low: ITT for OS, PFS and ORR (160 and 160 for each arm); SAS for AEs (158 and 147 for each arm); Total randomized patients: 160 and 160 for each arm | Low: All expected endpoints had been reported. | Low: Baseline characteristics were well balanced between both arms. |
| Fuchs 2018    | Unclear: No specific description    | Unclear: No specific description | Low: Double-blind placebo-controlled trial (participants and investigators) | Unclear: No specific description | Unclear: No specific description         | Unclear: No specific description         | Unclear: No specific description          |
| Matsuyama 2018| Unclear: No specific description    | Low: Dynamic allocation       | High: Open-label design               | Unclear: No specific description | Unclear: No specific description         | Low: All expected endpoints had been reported. | Unclear: No specific description          |
| Iqbal 2017    | Unclear: No specific description    | Unclear: No specific description | High: Open-label design               | Unclear: No specific description | Unclear: No specific description         | Unclear: No specific description         | Unclear: No specific description          |
| Year | Design | Randomization | Allocation | Blinding | Sample Size | Analysis | Reporting | Recruitment |
|------|--------|---------------|------------|----------|-------------|----------|-----------|-------------|
| Li 2017 | Low: Minimization method | Unclear: No specific description | High: Open-label design | Low: The reviews of all radiological scans were performed by two independent radiologists. | High: ITT for OS and PFS (71 and 74 for each arm); Measurable FAS for ORR (54 and 74 for each arm); SAS for AEs (71 and 74 for each arm); Total randomized patients: 71 and 74 for each arm | Low: All expected endpoints had been reported. | High: Early termination of patient recruitment |
| Hwang 2017 | Unclear: No specific description | Unclear: No specific description | High: Open-label design | Unclear: No specific description | Low: ITT for OS, PFS and ORR (24 and 26 for each arm); SAS for AEs (24 and 26 for each arm); Total randomized patients: 24 and 26 for each arm | Low: All expected endpoints had been reported. | High: Early termination of patient recruitment |
| Hall 2017 | Low: Permuted block randomization | Low: Central allocation by telephone | High: Treatment allocation was not masked. | Unclear: No specific description | Low: ITT for OS, PFS and ORR (17, 19 and 19 for each arm); SAS for AEs (17, 19 and 19 for each arm); Total randomized patients: 17, 19 and 19 for each arm | Low: All expected endpoints had been reported. | Low: Baseline characteristics were well balanced between both arms. |
| Li 2016 | Unclear: No specific description | Unclear: No specific description | High: It was virtually impossible to keep both arms blinded since their drug administrations were greatly different. | Unclear: No specific description | Unclear: No specific description | Unclear: No specific description | Unclear: No specific description |
| Yoon 2016 | Unclear: No specific description | Unclear: No specific description | Low: Double-blind placebo-controlled trial (participants and investigators) | Unclear: No specific description | Low: ITT for OS, PFS and ORR (84 and 84 for each arm); SAS for AEs (82 and 80 for each arm); Total | Low: All expected endpoints had been reported. | Low: Baseline characteristics were well balanced between both arms. |
| Study     | Methodology | Randomization | Blinding | Endpoint Assessment | Baseline Characteristics |
|-----------|-------------|---------------|----------|---------------------|--------------------------|
| Shah 2016 | Uncl. No specific description | Low: Allocation by interactive voice response system | Low: Double-blind placebo-controlled trial (participants and investigators) | Low: ORR (partial or complete best overall response) was assessed by investigators. | Low: All expected endpoints had been reported. |
| Tebbutt 2016 | Low: Minimization method | Uncl. No specific description | High: Open-label design | Unclear: No specific description | Low: All expected endpoints had been reported. |
| Hironaka 2016 | Low: Minimization method | Low: Central allocation by fax | High: Open-label design | Low: Disease progression and overall response were assessed by independent review committee. | Low: All expected endpoints had been reported. |
| Wang 2016 | Uncl. No specific description | Low: Central allocation | High: Open-label design | High: Tumor response was assessed by the investigators. | Low: All expected endpoints had been reported. |
| Study       | Allocation Method                           | Randomization  | Treatment                                          | Comparator                                          | Baseline Characteristics | Randomization Outcome |
|-------------|---------------------------------------------|----------------|----------------------------------------------------|----------------------------------------------------|--------------------------|------------------------|
| Du 2015     | Low: Permuted block randomization           | Low: Central allocation by interactive voice response system | High: Open-label design | Unclear: No specific description | SAS for AEs (119 and 115 for each arm); Total randomized patients: 121 and 122 for each arm | Low: All expected endpoints had been reported. |
| Wu 2015     | Low: Permuted block randomization           | Low: Allocation by sealed envelopes | High: Incomparable pattern of oral administrations between both arms | Unclear: No specific description | Low: ITT for OS, PFS and ORR (31 and 31 for each arm); SAS for AEs (31 and 31 for each arm); Total randomized patients: 32 and 32 for each arm | Unclear: No specific description |
| Van Cutsem 2015 | Unclear: No specific description             | Low: Central allocation by interactive voice response system | High: Open-label design | Unclear: No specific description | Low: ITT for OS, PFS and ORR (36 and 36 for each arm); SAS for AEs (36 and 36 for each arm); Total randomized patients: 36 and 36 for each arm | Low: Baseline characteristics were well balanced between both arms. |
| Shen 2015   | Low: Minimization method                    | Low: Allocation by interactive voice response system | Low: Double-blind placebo-controlled trial (participants and investigators) | Low: Tumor response was evaluated by investigators. | Low: FAS for OS, PFS and ORR (88, 81 and 78 for each arm); SAS for AEs (88, 81 and 78 for each arm); Total randomized patients: 89, 86 and 79 for each arm | Low: All expected endpoints had been reported. |

8
| Study | Randomization | Allocation | Design | Tumor Response | ITT endpoints | OS and PFS | ORR | AEs | Baseline characteristics |
|-------|---------------|------------|--------|----------------|---------------|-----------|-----|-----|-------------------------|
| Guimbau d 2014 | Unclear: No specific description | Unclear: No specific description | High: Open-label design | High: Tumor response was evaluated by investigators. | Low: ITT for OS and PFS (209 and 207 for each arm); Measurable FAS for ORR (189 and 198 for each arm); SAS for AEs (200 and 203 for each arm); Total randomized patients: 209 and 207 for each arm | Low: All expected endpoints had been reported. | Low: Baseline characteristics were well balanced between both arms. |
| Iveson 2014 | Low: Permuted block randomization | Low: Allocation by interactive voice response system | Low: Double-blind placebo-controlled trial (participants and investigators) | Low: Tumor response was evaluated by investigators. | Low: ITT for OS and PFS (82 and 39 for each arm); Measurable FAS for ORR (76 and 38 for each arm); SAS for AEs (81 and 39 for each arm); Total randomized patients: 82 and 39 for each arm | Low: All expected endpoints had been reported. | Low: Baseline characteristics were well balanced between both arms. |
| Zhang 2014 | Unclear: No specific description | Unclear: No specific description | High: Incomparable pattern of intravenous administrations between both arms | Unclear: No specific description | Low: ITT for OS, PFS and ORR (30 and 26 for each arm); SAS for AEs (30 and 26 for each arm); Total randomized patients: 30 and 26 for each arm | Low: All expected endpoints had been reported. | Unclear: There was lacking of description on planned amount of patient recruitment. |
| Lu 2014 | Low: Computerized randomization | Low: Allocation by sealed envelopes | High: Incomparable pattern of intravenous administrations between both arms | Unclear: No specific description | Low: ITT for OS, PFS and ORR (47 and 47 for each arm); SAS for AEs (47 and 47 for each arm); Total randomized patients: 47 and 47 for each arm | Low: All expected endpoints had been reported. | Low: Baseline characteristics were well balanced between both arms. |
| Study   | Randomization Method | Allocation | Label Design | Image Assessment | Endpoint Reporting | Other Observations |
|---------|----------------------|------------|--------------|-----------------|--------------------|-------------------|
| Sugimoto 2014 | Unclear: No specific description | Low: Central allocation | High: Open-label design | Low: All radiological assessments were confirmed by extratumoral review. | Low: All expected endpoints had been reported. | Baseline characteristics were well balanced between both arms. |
| Koizumi 2014 | Unclear: No specific description | Unclear: No specific description | High: Open-label design | High: Response rates were based on the assessment of response by the investigators at each center. | Low: All expected endpoints had been reported. | High: High percentage of patients (over 70%) in both arms received second-line regimens, which might dilute the difference of overall survival between both arms. |
| Koizumi 2013 | Low: Minimization method | Unclear: No specific description | High: Open-label design | Low: All measured images were assessed by a Central Imaging Review Committee. | Low: All expected endpoints had been reported. | High: High percentage of patients (over 90%) in both arms received second-line regimens, which might dilute the difference of overall survival between both arms. |
| Shirao 2013 | Low: Minimization method | Low: Central allocation | High: Open-label design | Low: No response-relevant endpoints had been reported while other endpoints were | Low: All expected endpoints had been reported. | High: High percentage of patients (over 70%) in both arms received second-line regimens, which might dilute the difference of overall survival between both arms. |
| Study       | Allocation Type | Randomization Type | Study Design | Endpoint Reporting | Baseline Characteristics |
|-------------|-----------------|--------------------|--------------|-------------------|--------------------------|
| Richards 2013 | Unclear: No specific description | Low: Central allocation | High: Open-label design | Unclear: No specific description | Low: ITT for OS and PFS (75 and 75 for each arm); Measurable FAS for ORR (68 and 71 for each arm); SAS for AEs (68 and 72 for each arm); Total randomized patients: 75 and 75 for each arm | Low: All expected endpoints had been reported. |
|             |                 | Low: Permuted block randomization | Low: Central allocation by fax | | Low: Baseline characteristics were well balanced between both arms. |
| Waddell 2013 | Low: Permuted block randomization | Low: Central allocation by fax | High: Open-label design | Unclear: No specific description | Low: ITT for OS and PFS (278 and 275 for each arm); Measurable FAS for ORR (254 and 238 for each arm); SAS for AEs (276 and 266 for each arm); Total randomized patients: 278 and 275 for each arm | Low: All expected endpoints had been reported. |
|             |                 | Low: Central allocation by interactive voice response system | High: Open-label design | Low: Date of progression and best overall response were assessed by masked review at an independent review committee | Low: Baseline characteristics were well balanced between both arms. |
| Lordick 2013 | Low: Permuted block randomization | Low: Central allocation by interactive voice response system | High: Open-label design | Low: ITT for OS, PFS and ORR (455 and 449 for each arm); SAS for AEs (446 and 436 for each arm); Total randomized patients: 455 and 449 for each arm | Low: All expected endpoints had been reported. |
|             |                 | Low: Permuted block randomization | High: Open-label design | | Low: Baseline characteristics were well balanced between both arms. |
| Study | Randomization Method | Allocation Method | Tumor Assessment | ITT/PFS/ORR | AEs | Total Patients | Baseline Characteristics |
|-------|----------------------|-------------------|-----------------|-------------|-----|---------------|--------------------------|
| Wang 2013 | Low: Computerized randomization | Low: Allocation by sealed envelopes | High: Tumor assessment was undertaken with CT or MRI consistently every 2 months by principal investigator. | Low: ITT for OS, PFS and ORR (41 and 41 for each arm); SAS for AEs (41 and 41 for each arm); Total randomized patients: 41 and 41 for each arm | Low: All expected endpoints had been reported. | Low: Baseline characteristics were well balanced between both arms. |
| Eatock 2013 | Low: Computerized randomization | Low: Allocation by interactive voice response system | Low: Radiologic tumour assessment of the chest, abdomen, and pelvis was performed by the investigators. | Low: ITT for OS and PFS (115 and 56 for each arm); Measurable FAS for ORR (100 and 49 for each arm); SAS for AEs (114 and 53 for each arm); Total randomized patients: 115 and 56 for each arm | Low: All expected endpoints had been reported. | Low: Baseline characteristics were well balanced between both arms. |
| Al-Batran 2013 | Unclear: No specific description | Unclear: No specific description | High: Open-label design | Unclear: No specific description | Low: All expected endpoints had been reported. | Low: Baseline characteristics were well balanced between both arms. |
| Andrić 2012 | Unclear: No specific description | Unclear: No specific description | Unclear: No specific description | Unclear: No specific description | Unclear: No specific description | Unclear: No specific description |
| Study     | Randomization Method | Allocation Method | Arm Administration | ITT & AEs | Overall Randomization Patients | Expected Endpoints | Baseline Characteristics |
|-----------|----------------------|-------------------|---------------------|----------|-------------------------------|---------------------|--------------------------|
| Roy 2012  | Unclear: No specific description | Unclear: No specific description | High: Incomparable pattern of intravenous administrations between both arms | Unclear: No specific description | Low: ITT for OS, ORR and AEs (42 and 43 for each arm); Total randomized patients: 43 and 43 for each arm | Low: All expected endpoints had been reported. | Low: Baseline characteristics were well balanced between both arms. |
| Mochiki 2012 | Unclear: No specific description | Low: Central allocation | High: Open-label design | Unclear: No specific description | Low: ITT for OS, PFS, ORR and AEs (42 and 41 for each arm); Total randomized patients: 42 and 41 for each arm | Low: All expected endpoints had been reported. | Low: Baseline characteristics were well balanced between both arms. |
| Ohtsu 2011 | Low: Permuted block randomization | Low: Allocation by interactive voice response system | Low: Double-blind placebo-controlled trial (participants and investigators) | Low: Although no independent radiologic review was performed, the double-blind design of this trial guaranteed the blinding of outcome assessment among investigators. | Low: ITT for OS and PFS (387 and 387 for each arm); Measurable FAS for ORR (311 and 297 for each arm); SAS for AEs (386 and 381 for each arm); Total randomized patients: 387 and 387 for each arm | Low: All expected endpoints had been reported. | Low: Baseline characteristics were well balanced between both arms. |
| Jeung 2011 | Unclear: No specific description | Unclear: No specific description | High: Incomparable pattern of oral administrations between both arms | Low: All responses were confirmed by a panel of independent intramural radiologists. | Low: ITT for OS, PFS, ORR and AEs (39 and 41 for each arm); Total randomized patients: 39 and 41 for each arm | Low: All expected endpoints had been reported. | Low: Baseline characteristics were well balanced between both arms. |
| Study            | Minimization Method | Centralization | Intravenous Administration | Responses | Randomization | Baseline Characteristics | Overall Survival |
|------------------|---------------------|----------------|-----------------------------|-----------|---------------|--------------------------|------------------|
| Komatsu 2011     | Low: Minimization method | Unclear: No specific description | High: Incomparable pattern of intravenous administrations between both arms | Unclear: No specific description | Low: ITT for OS, PFS, ORR and AEs (48 and 47 for each arm); Total randomized patients: 48 and 47 for each arm | Low: All expected endpoints had been reported. | High: The age ratio between both arms were not comparable. |
| Li 2011          | Unclear: No specific description | Unclear: No specific description | High: Incomparable pattern of intravenous administrations between both arms | Low: Responses were assessed by the independent review committee. | Low: ITT for OS, ORR and AEs (50 and 44 for each arm); Total randomized patients: 50 and 44 for each arm | Unclear: No specific description | Unclear: There was lacking of description on planned amount of patient recruitment. |
| Narahara 2011    | Unclear: No specific description | Low: Central dynamic allocation | High: Open-label design | Low: All radiologic assessments were confirmed by extramural review. | Low: FAS for OS and PFS (155 and 160 for each arm); Measurable FAS for ORR (94 and 93 for each arm); SAS for AEs (155 and 160 for each arm); Total randomized patients: 164 and 162 for each arm | Low: All expected endpoints had been reported. | High: High percentage of patients (over 70%) in both arms received second-line regimens, which might dilute the difference of overall survival between both arms. |
| Tebbutt 2010     | Unclear: No specific description | Low: Central allocation | High: Open-label design | Low: The tumour response of each patient was centrally reviewed by the lead study clinician and a clinician independent of the study. | Low: ITT for OS and PFS (50 and 56 for each arm); Measurable FAS for ORR (47 and 53 for each arm); SAS for AEs (49 and 55 for each arm); Total randomized patients: 50 and 56 for each arm | Low: All expected endpoints had been reported. | Low: Baseline characteristics were well balanced between both arms. |
| Study | Allocation | Randomization | Design | Endpoint Description | Baseline Characteristics |
|-------|------------|---------------|--------|----------------------|--------------------------|
| Yun 2010 | Unclear: No specific description | Unclear: No specific description | High: Open-label design | Unclear: No specific description | Low: All expected endpoints had been reported. |
|        | Low: ITT for PFS (44 and 47 for each arm); Measurable FAS for ORR (43 and 45 for each arm); SAS for AEs (44 and 45 for each arm); Total randomized patients: 44 and 47 for each arm |
|        | Low: Baseline characteristics were well balanced between both arms. |
| Moehler 2010 | Unclear: No specific description | Low: Central allocation | High: Open-label design | Unclear: No specific description | Low: All expected endpoints had been reported. |
|        | Low: FAS for OS and PFS (53 and 50 for each arm); Measurable FAS for ORR (53 and 50 for each arm); SAS for AEs (57 and 55 for each arm); Total randomized patients: 60 and 58 for each arm |
|        | Low: Baseline characteristics were well balanced between both arms. |
| Ikeda 2009 | Unclear: No specific description | Unclear: No specific description | High: Incomparable pattern of intravenous administrations between both arms | Unclear: No specific description | Low: All expected endpoints had been reported. |
|        | Unclear: No specific description |
| Boku 2009 | Low: Minimization method | Low: Central allocation | High: Open-label design | Low: CT and endoscopic images of responders taken every 2 months independently of the treatment schedule were reviewed | Low: All expected endpoints had been reported. |
|        | Low: ITT for OS and PFS (236, 234 and 234 for each arm); Measurable FAS for ORR (181, 174 and 175 for each arm); SAS for AEs (234, 234 and 232 for each arm); Total randomized patients: 236, 234 and 234 for each arm |
|        | High: High percentage of patients (over 70%) in both arms received second-line regimens, which might dilute the difference of overall survival between both arms. |
| Study     | Randomization   | Blinding                | Endpoint Reporting          | Baseline Characteristics |
|-----------|-----------------|-------------------------|-----------------------------|--------------------------|
| Ridwelski 2008 | Unclear: No specific description | Unclear: No specific description | High: Incomparable pattern of intravenous administrations between both arms | Unclear: No specific description | Low: All expected endpoints had been reported. | Unclear: No specific description |
| Tesselaar 2008 | Unclear: No specific description | Unclear: No specific description | High: Incomparable pattern of intravenous administrations between both arms | Unclear: No specific description | Low: All expected endpoints had been reported. | Unclear: No specific description |
| Jin 2008    | Unclear: No specific description | Unclear: No specific description | High: Open-label design | Unclear: No specific description | Low: All expected endpoints had been reported. | Unclear: No specific description |
| Dank 2008   | Low: Randomization by biased coin method | Unclear: No specific description | High: Incomparable pattern of intravenous administrations between both arms | Low: Used IRC (Independent Review Committee) evaluation as RR results | Low: All expected endpoints had been reported. | Low: Baseline characteristics were well balanced between both arms |
|Study | Randomization | Allocation | Design | Imaging | Endpoint Reporting | Baseline Characteristics |
|------|---------------|------------|--------|--------|-------------------|--------------------------|
| Koizumi 2008 | Low: Randomization by biased coin method | Low: Central allocation | High: Open-label design | Low: All images (CT, MRI, or chest radiography) were assessed by an extramural review committee. | Low: ITT for OS, PFS and AEs (148 and 150 for each arm); Measurable FAS for ORR (87 and 106 for each arm); Total randomized patients: 153 and 152 for each arm | Low: All expected endpoints had been reported. |
| Park 2008 | Unclear: No specific description | Unclear: No specific description | High: Open-label design | Unclear: No specific description | Low: ITT for OS, PFS and ORR (45 and 46 for each arm); SAS for AEs (45 and 45 for each arm); Total randomized patients: 45 and 46 for each arm | Low: All expected endpoints had been reported. |
| Popov 2008 | Unclear: No specific description | Unclear: No specific description | High: Incomparable pattern of intravenous administrations between both arms | Low: Independent response review was performed by members not | Low: ITT for OS and ORR (30 and 30 for each arm); Total randomized patients: 30 and 30 for each arm | Low: Baseline characteristics were well balanced between both arms. |
| Study          | Randomization Method | Allocation | Randomization Method | Baseline Characteristics | Expected Endpoints | ITT and FAS Endpoints | Randomized Patients |
|---------------|----------------------|------------|----------------------|--------------------------|--------------------|-----------------------|---------------------|
| Roth 2007     | Unclear: No specific description | Low: Central allocation | High: Incomparable pattern of intravenous administrations between both arms | Low: All responses were confirmed by an independent panel of radiologists and an oncologist. | Low: All expected endpoints had been reported. | Low: ITT for OS, PFS, ORR and AEs (41, 40 and 38 for each arm); Total randomized patients: 121 | |
| Lutz 2007     | Low: Minimization method | Low: Central allocation | High: Open-label design | Unclear: The data files and computed tomography scans were reviewed centrally by the study coordinators. | Low: All expected endpoints had been reported. | Low: ITT for OS and PFS (54, 54 and 37 for each arm); Measurable FAS for ORR (46, 48 and 33 for each arm); SAS for AEs (51, 53 and 37 for each arm); Total randomized patients: 54, 54 and 37 for each arm | 224 and 221 for each arm |
| Van Cutsem 2006 | Unclear: No specific description | Low: Central allocation | High: Incomparable pattern of intravenous administrations between both arms | Low: All radiologic assessments were reviewed by an external response review committee. | Low: All expected endpoints had been reported. | Low: FAS for OS, PFS and AEs (224 and 221 for each arm); Measurable FAS for ORR (224 and 221 for each arm); Total randomized patients: 230 and 227 for each arm | 230 and 227 for each arm |
| Ajani 2005    | Unclear: No specific description | Low: Central allocation | High: Open-label design | Low: All pertinent imaging studies were reviewed | Low: All expected endpoints had been reported. | Low: mITT for OS, PFS, ORR and AEs (79 and 76 for each arm); Total randomized patients: 158 | |

Baseline characteristics were well balanced between both arms.
| Study | Allocation | Randomization | Reporting | Baseline Characteristics |
|-------|------------|---------------|-----------|--------------------------|
| Moehler 2005 | Unclear: No specific description | Low: Central allocation | Unclear: No specific description | Low: All expected endpoints had been reported. | Low: Baseline characteristics were well balanced between both arms. |
| Thuss-Patience 2005 | Unclear: No specific description | Unclear: No specific description | High: Incomparable pattern of intravenous administrations between both arms | Low: Tumor response was assessed by an independent radiologist. | Low: Baseline characteristics were well balanced between both arms. |
| Pozzo 2004 | Low: Minimization method | Unclear: No specific description | High: Incomparable pattern of intravenous administrations between both arms | Low: An external response review committee reviewed radiological and clinical documentation for all patients in the study. | Low: Baseline characteristics were well balanced between both arms. |
| Bouché 2004 | Unclear: No specific description | Unclear: No specific description | High: Open-label design | Low: All expected endpoints had been reported. | Low: Baseline characteristics were well balanced between both arms. |
| Study          | Randomization Method | Allocation Method | Characteristics Description | Endpoints Evaluation | Randomization | Recruitment | Duration |
|---------------|----------------------|-------------------|----------------------------|----------------------|---------------|-------------|----------|
| Koizumi 2004  | Unclear: No specific description | Unclear: No specific description | High: Incomparable pattern of intravenous administrations between both arms | Low: All evaluations, including eligibility and response to the treatment, were performed by an independent evaluation committee. | Low: ITT for OS, ORR and AEs (32 and 29 for each arm); Total randomized patients: 32 and 29 for each arm | Unclear: No specific description | Unclear: There was lacking of description on planned amount of patient recruitment. |
| Cocconi 2003  | Unclear: No specific description | Low: Central allocation | High: Incomparable pattern of intravenous administrations between both arms | High: Response was assessed by the clinical investigators at each participating unit. | Low: FAS for OS (98 and 97 for each arm); Measurable FAS for ORR (98 and 97 for each arm); SAS for AEs (94 and 93 for each arm); Total randomized patients: 100 and 100 for each arm | Low: ITT for OS, PFS and ORR (70, 105 and 105 for each arm); SAS for AEs (67, 102 and 104 for each arm); Total randomized patients: 70, 105 and 105 for each arm | Low: All expected endpoints had been reported. | High: Early termination of patient recruitment |
| Ohtsu 2003    | Low: Minimization method | Unclear: No specific description | High: Incomparable pattern of intravenous administrations between both arms | Unclear: Objective responses were confirmed by central review at regular group meetings. | Low: FAS for ORR (98 and 97 for each arm); SAS for AEs (94 and 93 for each arm); Total randomized patients: 100 and 100 for each arm | Low: ITT for OS, PFS and ORR (70, 105 and 105 for each arm); SAS for AEs (67, 102 and 104 for each arm); Total randomized patients: 70, 105 and 105 for each arm | Low: All expected endpoints had been reported. | Low: Baseline characteristics were well balanced between both arms. |
| Tebbutt 2002  | Low: Computerized randomization | Low: Central allocation | High: Incomparable pattern of intravenous administrations between both arms | Unclear: No specific description | Low: ITT for OS and PFS (125 and 129 for each arm); Measurable FAS for ORR (118 and 121 for each arm); Total randomized patients: 125, 129 and 129 for each arm | Low: ITT for OS and PFS (125 and 129 for each arm); Measurable FAS for ORR (118 and 121 for each arm); Total randomized patients: 125, 129 and 129 for each arm | Low: All expected endpoints had been reported. | Low: Baseline characteristics were well balanced between both arms. |
| Study      | Low: Minimization method | Low: Central allocation | High: Incomparable pattern of intravenous administrations between both arms | Unclear: No specific description | SAS for AEs (123 and 127 for each arm); Total randomized patients: 125 and 129 for each arm | Low: Baseline characteristics were well balanced between both arms. |
|------------|--------------------------|-------------------------|---------------------------------------------------------------------------|-----------------------------------|----------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------|
| Kim 2001   | Low: Minimization method | Low: Central allocation | High: Incomparable pattern of intravenous administrations between both arms | Unclear: No specific description | Unlclear: No specific description | Low: All expected endpoints had been reported. |
| Vanhoefe r 2000 | Low: Minimization method | Low: Central allocation | High: Incomparable pattern of intravenous administrations between both arms | Unclear: No specific description | Low: ITT for OS and PFS (132, 134 and 133 for each arm); Measurable FAS for ORR (79, 81 and 85 for each arm); Total randomized patients: 132, 134 and 133 for each arm | Low: Baseline characteristics were well balanced between both arms. |
| Roth 1999  | Unclear: No specific description | Unclear: No specific description | High: Incomparable pattern of intravenous administrations between both arms | Unclear: No specific description | Unclear: No specific description | Unclear: No specific description |
| Waters 1999| Unclear: No specific description | Low: Central allocation by telephone | High: Incomparable pattern of intravenous administrations between both arms | Unclear: No specific description | Low: ITT for OS, PFS and AEs (130 and 126 for each arm); Measurable FAS for ORR (116 and 121 for each arm); Total randomized patients: 130 and 126 for each arm | Low: Baseline characteristics were well balanced between both arms. |
| Study       | Allocation Method | Randomization Method | Administration Pattern | Description | Analysis Plan | Recruitment | Notes                                      |
|-------------|-------------------|----------------------|-------------------------|-------------|---------------|-------------|--------------------------------------------|
| Icli 1998   | Unclear: No specific description | Low: Telephone allocation | High: Incomparable pattern of intravenous administrations between both arms | Unclear: No specific description | Low: ITT for OS, PFS and AEs (67 and 64 for each arm); Measurable FAS for ORR (59 and 59 for each arm); Total randomized patients: 67 and 64 for each arm | Low: All expected endpoints had been reported. | Unclear: There was lacking of description on planned amount of patient recruitment. |
| Yamamura 1998 | Unclear: No specific description | Unclear: No specific description | High: Incomparable pattern of intravenous administrations between both arms | Unclear: No specific description | Unclear: No specific description | Unclear: No specific description | Unclear: No specific description |
| Barone 1998 | Low: Permuted block randomization | Low: Sealed envelopes allocation | High: Incomparable pattern of intravenous administrations between both arms | Unclear: No specific description | Low: FAS for OS and ORR (33 and 32 for each arm); Total randomized patients: 36 and 36 for each arm | Low: All expected endpoints had been reported. | Unclear: There was lacking of description on planned amount of patient recruitment. |
| Scheithauer 1996 | Unclear: No specific description | Unclear: No specific description | High: Incomparable pattern of intravenous administrations between both arms | Unclear: No specific description | Unclear: No specific description | Unclear: No specific description | Unclear: No specific description |
| Colucci 1995 | Unclear: No specific description | Unclear: No specific description | High: Incomparable pattern of intravenous administrations between both arms | Unclear: No specific description | Unclear: No specific description | Unclear: No specific description | Unclear: No specific description |
| Pyrhönen 1995 | Low: Permuted block randomization | Low: Sealed envelopes allocation | High: Incomparable pattern of intravenous administrations between both arms | Unclear: No specific description | Low: ITT for OS, PFS, ORR and AEs (21 and 20 for each arm); Total randomized patients: 21 and 20 for each arm | Low: All expected endpoints had been reported. | High: Early termination of patient recruitment |
| Study | Allocation | Centralization | Intravenous Administration | Endpoints | Baseline Characteristics |
|-------|------------|----------------|-----------------------------|-----------|--------------------------|
| Coombes 1994 | Low: Allocation by telephone | High: Incomparable pattern of intravenous administrations between both arms | Low: Independent review was carried out to assess tumour response for all patients. | Low: ITT for OS, ORR and AEs (36 and 33 for each arm); Total randomized patients: 36 and 33 for each arm | Low: Baseline characteristics were well balanced between both arms. |
| Cocconi 1994 | Low: Central allocation | High: Incomparable pattern of intravenous administrations between both arms | Low: All patients were extramurally reviewed. | Low: mITT for OS (88 and 54 for each arm); Measurable FAS for ORR (85 and 52 for each arm); SAS for AEs (85 and 52 for each arm); Total randomized patients: 93 and 54 for each arm | Low: All expected endpoints had been reported. |
| Loehrer 1994 | Low: Central allocation | High: Incomparable pattern of intravenous administrations between both arms | Unclear: No specific description | Low: ITT for OS, PFS and AEs (69, 26 and 70 for each arm); Total randomized patients: 69, 26 and 70 for each arm | Low: All expected endpoints had been reported. |
| Cullinan 1994 | Unclear: No specific description | Unclear: No specific description | High: Incomparable pattern of intravenous administrations between both arms | Low: ITT for OS, PFS and AEs (79, 51, 53 and 69 for each arm); Total randomized patients: 79, 51, 53 and 69 for each arm | Unclear: No specific description |
| Study   | Randomization | Allocation | Design | Sample Size | Analysis | Early Termination |
|---------|---------------|------------|--------|-------------|----------|------------------|
| Murad 1993 | Unclear: No specific description | Unclear: No specific description | High: Incomparable pattern of intravenous administrations between both arms | Unclear: No specific description | Low: mITT for OS, ORR and AEs (30 and 10 for each arm); Total randomized patients: 30 and 10 for each arm | High: Early termination of patient recruitment |
| Kim 1993 | Unclear: No specific description | Unclear: No specific description | High: Incomparable pattern of intravenous administrations between both arms | Unclear: No specific description | Low: FAS for OS and AEs (98, 103 and 94 for each arm); Measurable FAS for ORR (57, 55 and 54 for each arm); Total randomized patients: 110, 112 and 102 for each arm | Low: All expected endpoints had been reported. |
| KRGGC 1992 | Unclear: No specific description | Unclear: No specific description | High: Incomparable pattern of intravenous administrations between both arms | Unclear: No specific description | Unclear: No specific description | Unclear: No specific description |
| Kelsen 1992 | Unclear: No specific description | Unclear: No specific description | High: Incomparable pattern of intravenous administrations between both arms | Unclear: No specific description | Low: ITT for OS and ORR (30 and 30 for each arm); Total randomized patients: 30 and 30 for each arm | High: Early termination of patient recruitment |
| Kikuchi 1990 | Unclear: No specific description | Low: Allocation by sealed envelopes | High: Incomparable pattern of intravenous administrations between both arms | Unclear: No specific description | Unclear: No specific description | Unclear: No specific description |
| GITSG 1988 | Unclear: No specific description | Unclear: No specific description | High: Incomparable pattern of intravenous administrations between both arms | Unclear: No specific description | Low: ITT for OS and AEs (85, 81 and 81 for each arm); Measurable FAS for ORR (30, 31 and 33 for each arm); Total randomized patients: 85, 81 and 81 for each arm | Unclear: No specific description |
| Name          | Date   | Allocation | Intravenous Admin. | ITT for OS | FAS for ORR and AEs | Total Randomized Patients | Endpoints | Patient Recruitment |
|---------------|--------|------------|--------------------|------------|---------------------|---------------------------|-----------|-------------------|
| Lacave        | 1987   | Low: by telephone | High: Incomparable | Unclear: No specific definition | Low: ITT for OS (85 and 88 for each arm); Measurable FAS for ORR (28 and 29 for each arm); Total randomized patients: 189 | Unclear: No specific description | Unclear: There was lacking of description on planned amount of patient recruitment. |
| Levi          | 1986   | Low: Central allocation | High: Incomparable | Unclear: No specific definition | Low: FAS for OS, PFS and AEs (94 and 93 for each arm); Measurable FAS for ORR (75 and 70 for each arm); Total randomized patients: 203 | Low: All expected endpoints had been reported. | Unclear: There was lacking of description on planned amount of patient recruitment. |
| De Lisi       | 1986   | Unclear: No specific description | High: Incomparable | Unclear: No specific definition | Unclear: No specific description | Unclear: No specific description | Unclear: No specific description |
| Cullinan      | 1985   | Unclear: No specific description | High: Incomparable | Unclear: No specific definition | Low: ITT for OS and PFS (51, 49 and 51 for each arm); Measurable FAS for ORR (11, 11 and 13 for each arm); Total randomized patients: 51, 49 and 51 for each arm | Low: All expected endpoints had been reported. | Low: Baseline characteristics were well balanced between both arms. |
| Douglass      | 1984   | Low: Central allocation | High: Incomparable | Unclear: No specific definition | Low: FAS for OS, ORR and AEs (39, 46, 44 and 46 for each arm); Total randomized patients: 52, 52, 54 and 52 for each arm | Unclear: No specific description | Low: Baseline characteristics were well balanced between both arms. |
| O’Connel      | 1984   | Unclear: No specific description | High: Incomparable | Unclear: No specific definition | Low: FAS for OS, ORR and AEs (76, 78 and 78 for each arm); Total randomized patients: 241 | Unclear: No specific description | Unclear: There was lacking of description on planned amount of patient recruitment. |
| Study  | Randomization | ITT for OS and AEs | FAS for ORR | Total randomized patients | Risk of Bias |
|--------|---------------|--------------------|-------------|---------------------------|--------------|
| Friedman 1983 | Permutated block randomization | Low: 36, 38, 34 and 34 for each arm | High: 22, 19, 12 and 22 for each arm | Unclear: No specific description | High: Incomparable pattern of intravenous administrations between both arms |
| O'Connel 1982 | Unclear: No specific description | Low: 43, 34, 58 and 46 for each arm | High: 12, 10, 19 and 18 for each arm | Unclear: No specific description | Unclear: No specific description |
| Buroker 1979 | Unclear: No specific description | Low: 80 and 87 for each arm | High: 43 and 54 for each arm | Unclear: No specific description | Unclear: No specific description |

**Note:** For selective reporting, if all planned endpoints of the protocol were reported in the article, it was undoubtedly regarded as low risk of bias, and vice versa. If there was no assessible protocol, studies that contained all five endpoints in our meta-analysis were also evaluated as low risk of bias, regardless of quantitative results or descriptive statements. Otherwise, unclear risk of bias was offered.

**Abbreviations:** ITT: intent-to-treat set; mITT: modified ITT; FAS: full-analysis set; PPS: per-protocol set; SAS: safety analysis set;
**eTable 3. Comparisons between direct and indirect results regarding overall survival (general analysis)**

| Comparison | Fixed (HR 95% CI) | Random (HR 95% CI) |
|------------|-------------------|--------------------|
|            | Direct            | Indirect           | Direct            | Indirect           |
| A1:F1      | 1.23 [0.80; 1.88] | 1.52 [1.15; 2.00]  | 1.23 [0.79; 1.90] | 1.52 [1.13; 2.06]  |
| A1:FA2     | 1.25 [0.73; 2.12] | 1.39 [1.06; 1.83]  | 1.25 [0.72; 2.16] | 1.38 [1.03; 1.86]  |
| A1:FA3     | 1.72 [1.28; 2.30] | 1.36 [0.97; 1.91]  | 1.72 [1.23; 2.39] | 1.35 [0.94; 1.92]  |
| F1:FA2     | 0.97 [0.72; 1.31] | 0.95 [0.80; 1.12]  | 0.97 [0.70; 1.34] | 0.94 [0.79; 1.13]  |
| F1:FA3     | 0.93 [0.76; 1.15] | 1.14 [1.02; 1.27]  | 0.94 [0.75; 1.17] | 1.13 [1.00; 1.28]  |
| F1:FE2     | 1.43 [0.86; 2.39] | 1.11 [0.88; 1.40]  | 1.43 [0.84; 2.44] | 1.10 [0.85; 1.43]  |
| F1:FI2     | 1.16 [0.95; 1.42] | 1.35 [1.19; 1.55]  | 1.17 [0.94; 1.46] | 1.37 [1.18; 1.59]  |
| F1:FP2     | 1.35 [1.20; 1.51] | 1.12 [1.00; 1.25]  | 1.36 [1.20; 1.54] | 1.13 [0.99; 1.28]  |
| F1:FP3     | 0.92 [0.65; 1.32] | 1.38 [1.26; 1.53]  | 0.93 [0.64; 1.35] | 1.41 [1.26; 1.58]  |
| F1:FT2     | 1.23 [1.06; 1.43] | 1.43 [1.22; 1.68]  | 1.27 [1.04; 1.54] | 1.45 [1.22; 1.72]  |
| F1:FY2     | 0.89 [0.74; 1.08] | 0.83 [0.63; 1.11]  | 0.89 [0.71; 1.11] | 0.85 [0.62; 1.16]  |
| F1:PA3     | 1.13 [0.74; 1.73] | 1.30 [0.89; 1.91]  | 1.13 [0.73; 1.76] | 1.30 [0.87; 1.94]  |
| F1:PI2     | 1.22 [1.00; 1.49] | 0.70 [0.47; 1.03]  | 1.22 [0.95; 1.56] | 0.70 [0.46; 1.07]  |
| FA2:FA3    | 1.12 [0.94; 1.34] | 1.17 [0.97; 1.42]  | 1.12 [0.93; 1.35] | 1.16 [0.94; 1.43]  |
| FA2:FP3    | 1.44 [1.12; 1.85] | 1.40 [1.19; 1.65]  | 1.45 [1.10; 1.91] | 1.43 [1.20; 1.71]  |
| FA2:S      | 0.46 [0.32; 0.67] | 0.37 [0.19; 0.74]  | 0.46 [0.31; 0.68] | 0.37 [0.19; 0.75]  |
| FA3:FE2    | 1.05 [0.81; 1.36] | 1.08 [0.77; 1.51]  | 1.05 [0.78; 1.41] | 1.10 [0.77; 1.58]  |
| FA3:FP2    | 1.07 [0.95; 1.21] | 1.16 [1.04; 1.29]  | 1.10 [0.93; 1.31] | 1.17 [1.03; 1.33]  |
| FA3:FP3    | 1.25 [1.12; 1.39] | 1.22 [1.09; 1.38]  | 1.27 [1.11; 1.45] | 1.26 [1.09; 1.45]  |
| FA3:PU2    | 0.98 [0.50; 1.24] | 0.68 [0.46; 1.01]  | 0.79 [0.50; 1.25] | 0.68 [0.44; 1.06]  |
| FA3:PA3    | 0.79 [0.43; 1.44] | 1.24 [0.90; 1.72]  | 0.79 [0.42; 1.47] | 1.26 [0.90; 1.78]  |
| FA3:S      | 0.33 [0.17; 0.64] | 0.40 [0.27; 0.60]  | 0.33 [0.17; 0.65] | 0.41 [0.27; 0.61]  |
| FE2:FI2    | 1.25 [0.83; 1.88] | 1.07 [0.83; 1.38]  | 1.25 [0.80; 1.93] | 1.08 [0.81; 1.44]  |
| FE2:FP2    | 0.96 [0.81; 1.13] | 0.94 [0.82; 1.07]  | 0.96 [0.79; 1.16] | 0.94 [0.81; 1.10]  |
| FI2:FP3    | 1.06 [0.88; 1.28] | 1.03 [0.90; 1.19]  | 1.07 [0.86; 1.33] | 1.04 [0.89; 1.21]  |
| FI2:FT2    | 1.01 [0.66; 1.55] | 1.03 [0.89; 1.19]  | 1.01 [0.64; 1.59] | 1.06 [0.89; 1.25]  |
| FI2:PI2    | 0.56 [0.39; 0.81] | 0.98 [0.78; 1.23]  | 0.56 [0.37; 0.84] | 0.97 [0.74; 1.28]  |
| FP3:FP2    | 0.88 [0.79; 0.98] | 0.94 [0.84; 1.04]  | 0.86 [0.75; 0.98] | 0.95 [0.84; 1.08]  |
| FP3:FT2    | 0.92 [0.67; 1.27] | 0.99 [0.87; 1.13]  | 0.92 [0.66; 1.29] | 1.02 [0.87; 1.19]  |
| FP3:PA3    | 1.23 [0.77; 1.97] | 0.76 [0.53; 1.09]  | 1.23 [0.75; 2.02] | 0.75 [0.51; 1.08]  |
| FP3:PT2    | 0.82 [0.67; 1.00] | 0.77 [0.61; 0.98]  | 0.82 [0.66; 1.01] | 0.76 [0.58; 0.99]  |
| FT2:FP2    | 0.56 [0.35; 0.90] | 0.87 [0.71; 1.05]  | 0.56 [0.34; 0.92] | 0.85 [0.68; 1.06]  |
| FT2:TI2    | 1.27 [0.83; 1.96] | 0.74 [0.54; 1.01]  | 1.27 [0.80; 2.01] | 0.72 [0.50; 1.03]  |
| FU2:FY2    | 1.16 [0.82; 1.65] | 1.00 [0.62; 1.63]  | 1.16 [0.79; 1.70] | 1.01 [0.61; 1.68]  |
| Study     | Random sequence generation | Allocation concealment | Blinding of participants and personnel | Blinding of outcome assessment | Incomplete outcome data | Selective reporting | Other sources of bias |
|-----------|----------------------------|------------------------|----------------------------------------|-------------------------------|------------------------|---------------------|-----------------------|
| Kawakami 2018 | Unclear: No specific description | Unclear: No specific description | High: Open-label design | High: Tumor response was assessed by investigators. | Low: ITT for OS, PFS and ORR (41 and 43 for each arm); SAS for AEs (39 and 43 for each arm); Total randomized patients: 41 and 43 for each arm | Low: All expected endpoints had been reported. | High: Early termination of patient recruitment |
| Nishikawa 2018 | Unclear: No specific description | Low: Central dynamic allocation | High: Open-label design | Unclear: No specific description | Low: mITT for OS and PFS (55 and 55 for each arm); Measurable FAS for ORR (36 and 33 for each arm); SAS for AEs (55 and 55 for each arm); Total randomized patients: 57 and 59 for each arm | Low: All expected endpoints had been reported. | High: Sex ratio was not comparable between both arms. |
| Ajani 2017 | Unclear: No specific description | Low: Allocation by interactive voice response system | High: Open-label design | High: Tumor response was assessed by investigators. | Low: ITT for OS and PFS (239 and 122 for each arm); Measurable FAS for ORR (193 and 91 for each arm); SAS for AEs (230 and 118 for each arm); Total randomized patients: 239 and 122 for each arm | Low: All expected endpoints had been reported. | High: Early termination of patient recruitment |
| Ryu 2016 | Unclear: No specific description | Unclear: No specific description | High: Open-label design | Unclear: No specific description | Low: FAS for OS, PFS, ORR and AEs (120 and 116 for each arm); Total randomized patients: 127 and 128 for each arm | Low: All expected endpoints had been reported. | Unclear: No specific description |
| Li 2015 | Unclear: No specific description | Low: Central allocation | High: Open-label design | High: Investigators assessed tumor | Low: Baseline characteristics were well balanced between both arms. | Low: All expected endpoints | Low: Baseline characteristics were well balanced between both arms. |
| Study      | Low: Permuted block randomization | Unclear: No specific description | High: Open-label design | Unclear: No specific description | Low: ITT for OS and PFS (29 and 27 for each arm); SAS for AEs (29 and 26 for each arm); Total randomized patients: 29 and 27 for each arm | had been reported. |
|------------|----------------------------------|----------------------------------|------------------------|----------------------------------|-------------------------------------------------------------------------------------------------|---------------------|
| Ochenduszk o 2015 | Low: Minimization method | Low: Central allocation | High: Open-label design | Low: All images for PFS and tumor responses were reviewed by an independent review committee. | Low: PPS for OS, PFS and ORR (318 and 324 for each arm); SAS for AEs (338 and 335 for each arm); Total randomized patients: 343 and 342 for each arm | Low: All expected endpoints had been reported. |
| Yamada 2015 | Low: Minimization method | Low: Central allocation | High: Open-label design | Low: All images for PFS and tumor responses were reviewed by an independent review committee. | Low: PPS for OS, PFS and ORR (318 and 324 for each arm); SAS for AEs (338 and 335 for each arm); Total randomized patients: 343 and 342 for each arm | High: Early termination of patient recruitment |
| Chen 2015 | Unclear: No specific description | Unclear: No specific description | High: Incomparable pattern of intravenous administrations between both arms | Unclear: No specific description | Unclear: No specific description | Unclear: No specific description |
| Li 2014 | Unclear: No specific description | Unclear: No specific description | Unclear: No specific description | Unclear: No specific description | Unclear: No specific description | Unclear: No specific description |
| Kim 2012 | Unclear: No specific description | Unclear: No specific description | High: Open-label design | Unclear: No specific description | Low: mITT for OS, PFS and AEs (65 and 64 for each arm); Measurable FAS for ORR (53 and 45 for each arm) | Low: All expected endpoints |
| Low: All expected endpoints had been reported. |
| Unclear: No specific description |
| Unclear: No specific description |
| Unclear: No specific description |
| Unclear: No specific description |
| Low: All expected endpoints |
| Low: Baseline characteristics were well matched. |
| Study       | Randomization Method | Allocation Method | Blinding Method | Review of Response | Data Reported | Baseline Characteristics Balanced |
|-------------|----------------------|-------------------|-----------------|-------------------|---------------|-----------------------------------|
| Ocirk 2012  | Unclear: No specific description | Low: Central allocation | High: Open-label design | Unclear: No specific description | Low: ITT for OS, PFS, ORR and AEs (45 and 40 for each arm); Total randomized patients: 65 and 65 for each arm | Low: All expected endpoints had been reported. |
|             |                      |                   |                 |                   | Low: All expected endpoints had been reported. | Low: Baseline characteristics were well balanced between both arms. |
| Ajani 2010  | Unclear: No specific description | Unclear: No specific description | High: Open-label design | Low: Radiographic evidence of response to treatment was independently reviewed. | Low: FAS for OS and PFS (521 and 508 for each arm); Measurable FAS for ORR (402 and 385 for each arm); SAS for AEs (521 and 508 for each arm); Total randomized patients: 527 and 526 for each arm | Low: All expected endpoints had been reported. |
|             |                      |                   |                 |                   | Low: All expected endpoints had been reported. | Low: Baseline characteristics were well balanced between both arms. |
| Lee 2009    | Low: Permuted block randomization | Low: Central allocation | Unclear: No specific description | High: Objective responses were confirmed by participating investigators. | Low: FAS for OS and PFS (88 and 86 for each arm); Measurable FAS for ORR (78 and 78 for each arm); SAS for AEs (88 and 86 for each arm); Total randomized patients: 184 | Low: All expected endpoints had been reported. |
|             |                      |                   |                 |                   | Low: All expected endpoints had been reported. | Low: Baseline characteristics were well balanced between both arms. |
| Kang 2009   | Low: Permuted block randomization | Unclear: No specific description | High: Open-label design | Low: An independent review committee (IRC) reviewed patients’ radiological images and | Low: ITT for OS and PFS (160 and 156 for each arm); Measurable FAS for ORR (139 and 137 for each arm); SAS for AEs (156 and 155 for each arm); Total randomized patients: 160 and 156 for each arm | Low: All expected endpoints had been reported. |
|             |                      |                   |                 |                   | Low: All expected endpoints had been reported. | Low: Baseline characteristics were well balanced between both arms. |
| Study          | Randomization Method | Allocation Method | Baseline Characteristics | ITT and FAS Description | Expected Endpoints | Randomized Patients | Risk of Bias |
|---------------|----------------------|-------------------|--------------------------|--------------------------|---------------------|---------------------|--------------|
| Popov 2008    | Unclear: No specific description | Unclear: No specific description | Unclear: No specific description | Low: Independent response review was performed by members not involved in the study. | Low: ITT for OS, PFS and ORR (36 and 36 for each arm); Total randomized patients: 36 and 36 for each arm | Low: All expected endpoints had been reported. | Low: Baseline characteristics were well balanced between both arms. |
| Al-Batran 2008 | Unclear: No specific description | Unclear: No specific description | Unclear: No specific description | Unclear: No specific description | Low: ITT for OS and PFS (112 and 108 for each arm); Measurable FAS for ORR (112 and 106 for each arm); SAS for AEs (112 and 102 for each arm); Total randomized patients: 112 and 108 for each arm | Low: All expected endpoints had been reported. | Low: Baseline characteristics were well balanced between both arms. |
| Cunningham 2008 | Low: Permutated block randomization | Low: Central allocation by telephone | High: Incomparable pattern of intravenous administrations between both arms | Unclear: No specific description | Low: ITT for OS, PFS and ORR (263, 250, 245 and 244 for each arm); SAS for AEs (234, 234, 225 and 227 for each arm); Total randomized patients: 263, 250, 245 and 244 for each arm | Low: All expected endpoints had been reported. | Low: Baseline characteristics were well balanced between both arms. |

**Note:** Studies from general analysis that were also eligible for additional analysis were not repeatedly listed here. For selective reporting, if all planned endpoints of the protocol were reported in the article, it was undoubtedly regarded as low risk of bias, and vice versa. If there was no assessible
protocol, studies that contained all five endpoints in our meta-analysis were also evaluated as low risk of bias, regardless of quantitative results or descriptive statements. Otherwise, unclear risk of bias was offered.

**Abbreviations:** ITT: intent-to-treat set; mITT: modified ITT; FAS: full-analysis set; PPS: per-protocol set; SAS: safety analysis set;
| Study       | Random sequence generation | Allocation concealment | Blinding of participants and personnel | Blinding of outcome assessment | Incomplete outcome data                                                                 | Selective reporting | Other sources of bias                          |
|-------------|----------------------------|------------------------|----------------------------------------|-------------------------------|-----------------------------------------------------------------------------------------|---------------------|-----------------------------------------------|
| Tabernero 2018 | Low: Permuted block randomization | Low: Allocation by interactive voice response system | Low: Double-blind placebo-controlled trial (participants and investigators) | Low: Tumor response was assessed by investigators. | Low: ITT for OS and PFS (388 and 392 for each arm); Measurable FAS for ORR (351 and 352 for each arm); SAS for AEs (385 and 388 for each arm); Total randomized patients: 388 and 392 for each arm | Low: All expected endpoints had been reported. | Low: Baseline characteristics were well balanced between both arms. |
| Moehler 2018 | Unclear: No specific description | Unclear: No specific description | Low: Double-blind placebo-controlled trial (participants and investigators) | Unclear: No specific description | Low: mITT for OS, PFS, ORR and AEs (14 and 14 for each arm); Total randomized patients: 15 and 14 for each arm | Low: All expected endpoints had been reported. | High: Early termination of patient recruitment |
| Hecht 2016  | Unclear: No specific description | Low: Central allocation | Low: Quadruple-blind placebo-controlled trial (participant, care provider, investigator, outcomes assessor) | Low: Quadruple-blind placebo-controlled trial (participant, care provider, investigator, outcomes assessor) | Low: FAS for OS, PFS and ORR (249 and 238 for each arm); SAS for AEs (270 and 267 for each arm); Total randomized patients: 272 and 273 for each arm | Low: All expected endpoints had been reported. | Low: Baseline characteristics were well balanced between both arms. |
| Bang 2010   | Low: Permuted block randomization | Low: Allocation by interactive voice | High: Open-label design | High: Tumor response was not independently assessed. | Low: FAS for OS, PFS, ORR and AEs (294 and 290 for each arm); Total randomized patients: 298 and 296 for each arm | Low: All expected endpoints had been reported. | Low: Baseline characteristics were well balanced between both arms. |
| Study          | Randomization Method            | Response System                               | Endpoint Details                                                                 | Baseline Characteristics                                                | Comment                                      |
|---------------|---------------------------------|-----------------------------------------------|----------------------------------------------------------------------------------|--------------------------------------------------------------------------|----------------------------------------------|
| Catenacci 2017 | Low: Permuted block randomization | Low: Allocation by interactive voice response system | Low: Triple-blind placebo-controlled trial (participant, care provider, investigator) | Unclear: No specific description                                         | Low: All expected endpoints had been reported. |
|               |                                 |                                               | Low: ITT for OS and PFS (304 and 305 for each arm); Measurable FAS for ORR (262 and 267 for each arm); SAS for AEs (298 and 299 for each arm); Total randomized patients: 304 and 305 for each arm | Low: Baseline characteristics were well balanced between both arms.     |                                              |
| Shah 2017     | Low: Permuted block randomization | Unclear: No specific description               | Low: Double-blind placebo-controlled trial (participants and investigators)       | Low: Tumor response was assessed by investigators.                        | Low: All expected endpoints had been reported. | High: Early termination of patient recruitment |
|               |                                 |                                               | Low: ITT for OS and PFS (279 and 283 for each arm); Measurable FAS for ORR (207 and 217 for each arm); SAS for AEs (279 and 280 for each arm); Total randomized patients: 279 and 283 for each arm |                                              |                                              |
| Schuler 2016  | Unclear: No specific description | Unclear: No specific description               | High: Open-label design                                                             | Unclear: No specific description                                         | Low: All expected endpoints had been reported. |                                              |
| Rao 2010      | Low: Computerized randomization  | Low: Allocation by interactive voice response system | Low: Blinded radiological review was performed by an IRC in addition to local institute assessment. | Low: ITT for OS, PFS, ORR and AEs (35 and 36 for each arm); Total randomized patients: 35 and 36 for each arm | Low: Baseline characteristics were well balanced between both arms. |                                              |

**Note:** For selective reporting, if all planned endpoints of the protocol were reported in the article, it was undoubtedly regarded as low risk of bias, and vice versa. If there was no assessible protocol, studies that contained all five endpoints in our meta-analysis were also evaluated as low risk of bias, regardless of quantitative results or descriptive statements. Otherwise, unclear risk of bias was offered.

**Abbreviations:** ITT: intent-to-treat set; mITT: modified ITT; FAS: full-analysis set; PPS: per-protocol set; SAS: safety analysis set;
Figure 1. Funnel plot of overall survival (general analysis)
eFigure 2. Network forest plot for progression-free survival (general analysis)

| Treatment | (Random Effects Model) | HR   | 95%-CI   | P-Score |
|-----------|------------------------|------|----------|---------|
|            |                        | 0.75 | [0.54; 1.04] | 0.919   |
| FP3R      |                        | 0.83 | [0.71; 0.96] | 0.881   |
| FP3       |                        | 0.77 | [0.37; 1.59] | 0.824   |
| PA3       |                        | 0.91 | [0.77; 1.08] | 0.780   |
| Fl2       |                        | 0.91 | [0.75; 1.10] | 0.778   |
| FT2       |                        | 0.97 | [0.84; 1.13] | 0.702   |
| FP2R      |                        | 0.97 | [0.57; 1.65] | 0.687   |
| FA4       |                        | 1.00 |          | 0.662   |
| FP2       |                        | 1.25 | [0.70; 2.25] | 0.460   |
| PT2R      |                        | 1.25 | [0.98; 1.59] | 0.441   |
| PT2       |                        | 1.32 | [0.63; 2.78] | 0.430   |
| P1        |                        | 1.27 | [1.01; 1.61] | 0.428   |
| FE2       |                        | 1.30 | [0.78; 2.15] | 0.422   |
| Pi2       |                        | 1.35 | [0.95; 1.92] | 0.371   |
| FA2       |                        | 1.37 | [0.97; 1.93] | 0.358   |
| TI2       |                        | 1.43 | [0.90; 2.29] | 0.336   |
| F1        |                        | 1.48 | [1.28; 1.71] | 0.272   |
| FY2       |                        | 1.67 | [1.24; 2.26] | 0.195   |
| A1        |                        | 3.70 | [2.28; 6.02] | 0.039   |
| S         |                        | 4.48 | [2.55; 7.86] | 0.016   |
**eFigure 3. Network forest plot for objective response rate (general analysis)**

| Treatment | Comparison: other vs 'FP2' | RR       | 95%-CI      | P-Score |
|-----------|----------------------------|----------|-------------|---------|
| FP3R      |                            | 1.48 [1.11; 1.98] | 0.036 |
| FP3       |                            | 1.20 [1.06; 1.36] | 0.143 |
| FI2       |                            | 1.19 [1.00; 1.41] | 0.153 |
| FT2       |                            | 1.11 [0.94; 1.31] | 0.228 |
| FP2R      |                            | 1.10 [0.97; 1.25] | 0.232 |
| Pi2       |                            | 1.10 [0.80; 1.50] | 0.243 |
| PT2R      |                            | 1.10 [0.60; 2.04] | 0.259 |
| FP2       |                            | 1.00      |             | 0.347  |
| FA4       |                            | 0.90 [0.34; 2.42] | 0.410 |
| PA3       |                            | 0.88 [0.54; 1.45] | 0.426 |
| FA2       |                            | 0.83 [0.56; 1.24] | 0.476 |
| TI2       |                            | 0.82 [0.54; 1.26] | 0.486 |
| P1        |                            | 0.79 [0.45; 1.39] | 0.510 |
| PT2       |                            | 0.77 [0.60; 0.98] | 0.546 |
| FA3       |                            | 0.71 [0.57; 0.89] | 0.598 |
| FU3       |                            | 0.61 [0.25; 1.47] | 0.647 |
| FE2       |                            | 0.64 [0.43; 0.95] | 0.667 |
| AY2       |                            | 0.59 [0.36; 0.95] | 0.715 |
| F1        |                            | 0.60 [0.52; 0.69] | 0.728 |
| FY2       |                            | 0.48 [0.31; 0.73] | 0.815 |
| A1        |                            | 0.29 [0.15; 0.54] | 0.919 |
| FU2       |                            | 0.27 [0.16; 0.44] | 0.935 |
| S         |                            | 0.07 [0.01; 0.48] | 0.984 |

**Note:** The actual P-score of each regimen should be 1 minus its value displayed in the figure, according to the network calculation in R software.
eFigure 4. Network forest plot for hematological adverse events (general analysis)

| Treatment | Comparison: other vs 'FP2' | RR   | 95%-CI   | P-Score |
|-----------|-----------------------------|------|----------|---------|
| S         |                             | 0.16 | [0.02; 1.26] | 0.952   |
| A1        |                             | 0.34 | [0.19; 0.63] | 0.949   |
| F1        |                             | 0.39 | [0.32; 0.48] | 0.934   |
| F12       |                             | 0.75 | [0.59; 0.96] | 0.818   |
| P1        |                             | 0.76 | [0.36; 1.61] | 0.743   |
| FP2R      |                             | 0.94 | [0.74; 1.19] | 0.684   |
| FY2       |                             | 0.97 | [0.61; 1.53] | 0.643   |
| FA2       |                             | 0.99 | [0.67; 1.47] | 0.637   |
| FP2       |                             | 1.00 |           |         |
| FA4       |                             | 1.04 | [0.67; 1.64] | 0.591   |
| FT2       |                             | 1.12 | [0.83; 1.51] | 0.536   |
| FE2       |                             | 1.14 | [0.75; 1.74] | 0.520   |
| AY2       |                             | 1.20 | [0.65; 2.19] | 0.485   |
| FA3       |                             | 1.28 | [0.98; 1.68] | 0.419   |
| FP3R      |                             | 1.31 | [0.75; 2.29] | 0.414   |
| Ti2       |                             | 1.34 | [0.63; 2.85] | 0.407   |
| FU2       |                             | 1.50 | [0.92; 2.44] | 0.311   |
| FP3       |                             | 1.55 | [1.25; 1.90] | 0.272   |
| P12       |                             | 1.72 | [1.10; 2.71] | 0.220   |
| PT2R      |                             | 1.91 | [0.90; 4.06] | 0.201   |
| PA3       |                             | 2.43 | [0.60; 9.81] | 0.186   |
| PT2       |                             | 1.85 | [1.31; 2.60] | 0.167   |
| FU3       |                             | 2.01 | [1.06; 3.81] | 0.158   |
| FM2       |                             | 2.45 | [1.04; 5.81] | 0.117   |
eFigure 5. Network forest plot for non-hematological adverse events (general analysis)
eFigure 6. Network forest plot for overall survival (additional analysis)

| Treatment | Comparison: other vs 'FC2' (Random Effects Model) | HR | 95%-CI | P-Score |
|-----------|---------------------------------------------------|----|--------|---------|
| XC3R      |                                                   | 0.54 [0.34; 0.84] | 0.921  |
| FO3       |                                                   | 0.67 [0.54; 0.83] | 0.802  |
| XO3       |                                                   | 0.68 [0.54; 0.85] | 0.776  |
| SO2R      |                                                   | 0.61 [0.34; 1.11] | 0.772  |
| XC3       |                                                   | 0.68 [0.54; 0.87] | 0.755  |
| SO3       |                                                   | 0.73 [0.56; 0.95] | 0.651  |
| XO2       |                                                   | 0.74 [0.50; 1.10] | 0.610  |
| FC3       |                                                   | 0.75 [0.64; 0.86] | 0.604  |
| FO2       |                                                   | 0.77 [0.65; 0.92] | 0.574  |
| SO2       |                                                   | 0.83 [0.70; 0.99] | 0.450  |
| FH2       |                                                   | 0.83 [0.62; 1.11] | 0.449  |
| FO2R      |                                                   | 0.83 [0.58; 1.18] | 0.444  |
| XC2R      |                                                   | 0.88 [0.70; 1.11] | 0.356  |
| XO3R      |                                                   | 0.93 [0.66; 1.30] | 0.269  |
| XC2       |                                                   | 0.92 [0.74; 1.13] | 0.258  |
| SC3       |                                                   | 0.93 [0.76; 1.12] | 0.249  |
| SC2R      |                                                   | 0.96 [0.65; 1.42] | 0.239  |
| SC2       |                                                   | 0.93 [0.84; 1.04] | 0.220  |
| FC2       |                                                   | 1.00               | 0.102  |

Note: “FC2” was the common comparator.
**Note:** Treatments are hierarchically ranked according to their P-score. The higher position in the table a regimen locates at, the better survival benefits it could offer. Values situated at the intersection of specific column and row are the network effect sizes (HR and 95% CI) of row-defining regimen versus column-defining regimen.
**eFigure 8. Network forest plot for progression-free survival (additional analysis)**

| Treatment | Comparison: other vs 'FC2' (Random Effects Model) | HR  | 95%-CI | P-Score |
|-----------|--------------------------------------------------|-----|--------|---------|
| XC3R      |                                                  | 0.40[0.26; 0.61] | 0.981 |
| FO3       |                                                  | 0.59[0.45; 0.77] | 0.801 |
| XC3       |                                                  | 0.58[0.43; 0.79] | 0.799 |
| SO2R      |                                                  | 0.61[0.46; 0.81] | 0.763 |
| XO3       |                                                  | 0.62[0.46; 0.84] | 0.723 |
| SO3       |                                                  | 0.64[0.52; 0.79] | 0.687 |
| FC3       |                                                  | 0.74[0.57; 0.96] | 0.526 |
| FO2       |                                                  | 0.74[0.50; 1.11] | 0.509 |
| XO3R      |                                                  | 0.76[0.49; 1.17] | 0.502 |
| XO2       |                                                  | 0.76[0.51; 1.12] | 0.498 |
| FO2R      |                                                  | 0.81[0.61; 1.08] | 0.435 |
| XC2R      |                                                  | 0.81[0.65; 1.03] | 0.423 |
| SO2       |                                                  | 0.91[0.69; 1.20] | 0.285 |
| SC3       |                                                  | 0.92[0.79; 1.07] | 0.260 |
| SC2       |                                                  | 0.93[0.73; 1.18] | 0.253 |
| XC2       |                                                  | 1.00            | 0.160 |
| FC2       |                                                  | 1.00            | 0.160 |
| FH2       |                                                  | 1.22[0.81; 1.85] | 0.087 |
| SC2R      |                                                  | 1.44[0.93; 2.23] | 0.027 |

**Note:** “FC2” was the common comparator.
**Note:** Treatments are hierarchically ranked according to their P-score. The higher position in the table a regimen locates at, the better survival benefits it could offer. Values situated at the intersection of specific column and row are the network effect sizes (HR and 95% CI) of row-defining regimen versus column-defining regimen.
eFigure 10. Network forest plot for objective response rate (additional analysis)

| Treatment | Comparison: other vs 'FC2' (Random Effects Model) | RR  | 95%-CI   | P-Score |
|-----------|---------------------------------------------------|-----|----------|---------|
| XC3R      |                                                   | 2.23 [1.21; 4.11] | 0.083   |
| FO3       |                                                   | 1.76 [1.23; 2.52] | 0.188   |
| XC2R      |                                                   | 1.68 [1.15; 2.47] | 0.230   |
| SO2R      |                                                   | 1.56 [0.75; 3.22] | 0.367   |
| XC2       |                                                   | 1.48 [1.07; 2.05] | 0.379   |
| SO3       |                                                   | 1.52 [0.79; 2.92] | 0.387   |
| XO2       |                                                   | 1.48 [0.84; 2.61] | 0.386   |
| XC3       |                                                   | 1.46 [0.96; 2.22] | 0.403   |
| XO3R      |                                                   | 1.40 [0.81; 2.41] | 0.458   |
| FC3       |                                                   | 1.33 [1.04; 1.69] | 0.514   |
| XO3       |                                                   | 1.29 [0.88; 1.89] | 0.563   |
| FO2R      |                                                   | 1.24 [0.79; 1.96] | 0.583   |
| SC3       |                                                   | 1.24 [0.81; 1.91] | 0.596   |
| FO2       |                                                   | 1.23 [0.90; 1.67] | 0.622   |
| SG2R      |                                                   | 1.21 [0.79; 1.87] | 0.625   |
| SC2       |                                                   | 1.17 [0.92; 1.50] | 0.686   |
| SO2       |                                                   | 1.16 [0.83; 1.64] | 0.691   |
| FH2       |                                                   | 0.96 [0.66; 1.66] | 0.822   |
| FC2       |                                                   | 1.00               | 0.886   |

**Note:** The actual P-score of each regimen should be 1 minus its value displayed in the figure, according to the network calculation in R software. “FC2” was the common comparator.
### eFigure 11. Network league table for objective response rate (additional analysis)

| FC2  | FC4  | FC6  | FC8  | SC6  | SC8  | SC10 | SC12 |
|------|------|------|------|------|------|------|------|
| 0.68| 0.66| 0.48| 0.39| 0.66| 0.48| 0.39| 0.66|
| 1.10| 1.02| 0.80| 0.68| 1.10| 1.02| 0.80| 1.10|
| 0.17| 0.15| 0.13| 0.12| 0.17| 0.15| 0.13| 0.17|
| 0.27| 0.26| 0.25| 0.24| 0.27| 0.26| 0.25| 0.27|
| 1.23| 1.09| 0.93| 0.80| 1.23| 1.09| 0.93| 1.23|
| 2.28| 1.90| 1.61| 1.32| 2.28| 1.90| 1.61| 2.28|
| 2.28| 1.90| 1.61| 1.32| 2.28| 1.90| 1.61| 2.28|
| 1.23| 1.09| 0.93| 0.80| 1.23| 1.09| 0.93| 1.23|
| 1.23| 1.09| 0.93| 0.80| 1.23| 1.09| 0.93| 1.23|
| 0.40| 0.34| 0.29| 0.24| 0.40| 0.34| 0.29| 0.40|
| 0.40| 0.34| 0.29| 0.24| 0.40| 0.34| 0.29| 0.40|
| 0.34| 0.29| 0.24| 0.20| 0.34| 0.29| 0.24| 0.34|
| 0.34| 0.29| 0.24| 0.20| 0.34| 0.29| 0.24| 0.34|
| 0.20| 0.16| 0.12| 0.09| 0.20| 0.16| 0.12| 0.20|
| 0.20| 0.16| 0.12| 0.09| 0.20| 0.16| 0.12| 0.20|
| 0.16| 0.12| 0.09| 0.06| 0.16| 0.12| 0.09| 0.16|
| 0.16| 0.12| 0.09| 0.06| 0.16| 0.12| 0.09| 0.16|
| 0.06| 0.05| 0.04| 0.03| 0.06| 0.05| 0.04| 0.06|
| 0.06| 0.05| 0.04| 0.03| 0.06| 0.05| 0.04| 0.06|
| 0.05| 0.04| 0.03| 0.02| 0.05| 0.04| 0.03| 0.05|
| 0.05| 0.04| 0.03| 0.02| 0.05| 0.04| 0.03| 0.05|
| 0.03| 0.02| 0.01| 0.00| 0.03| 0.02| 0.01| 0.03|
| 0.03| 0.02| 0.01| 0.00| 0.03| 0.02| 0.01| 0.03|
| 0.01| 0.00| 0.00| 0.00| 0.01| 0.00| 0.00| 0.01|
| 0.01| 0.00| 0.00| 0.00| 0.01| 0.00| 0.00| 0.01|
| 0.00| 0.00| 0.00| 0.00| 0.00| 0.00| 0.00| 0.00|
| 0.00| 0.00| 0.00| 0.00| 0.00| 0.00| 0.00| 0.00|

**Note:** Treatments are hierarchically ranked according to their P-score. The lower position in the table a regimen locates at, the better response benefits it could offer. Values situated at the intersection of specific column and row are the network effect sizes (RR and 95%CI) of row-defining regimen versus column-defining regimen.
Note: “FC2” was the common comparator.
### eFigure 13. Network league table for hematological adverse events (additional analysis)

| Regimen | Note: Treatments are hierarchically ranked according to their P-score. The higher position in the table a regimen locates at, the better safety benefits it could offer. Values situated at the intersection of specific column and row are the network effect sizes (RR and 95%CI) of row-defining regimen versus column-defining regimen. |
eFigure 14. Network forest plot for non-hematological adverse events (additional analysis)

Note: “FC2” was the common comparator.
**eFigure 15. Network league table for non-hematological adverse events (additional analysis)**

| Treatment 1 | Treatment 2 | Treatment 3 | Treatment 4 | Treatment 5 | Treatment 6 | Treatment 7 | Treatment 8 | Treatment 9 | Treatment 10 |
|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| Treatment A |Treatment B  |Treatment C  |Treatment D  |Treatment E  |Treatment F  |Treatment G  |Treatment H  |Treatment I  |Treatment J  |
| 0.65        | 0.58        | 0.62        | 0.71        | 0.69        | 0.74        | 0.68        | 0.72        | 0.70        | 0.67        |
| 0.56        | 0.60        | 0.59        | 0.65        | 0.63        | 0.62        | 0.61        | 0.64        | 0.66        | 0.62        |
| 0.72        | 0.73        | 0.71        | 0.70        | 0.75        | 0.74        | 0.73        | 0.72        | 0.71        | 0.70        |
| 0.80        | 0.79        | 0.81        | 0.82        | 0.83        | 0.84        | 0.85        | 0.86        | 0.87        | 0.88        |
| 0.90        | 0.91        | 0.92        | 0.93        | 0.94        | 0.95        | 0.96        | 0.97        | 0.98        | 0.99        |

**Note:** Treatments are hierarchically ranked according to their P-score. The higher position in the table a regimen locates at, the better safety benefits it could offer. Values situated at the intersection of specific column and row are the network effect sizes (RR and 95%CI) of row-defining regimen versus column-defining regimen.
Supplementary figure legends

**eFigure 1.** Funnel plot of overall survival (general analysis)
**eFigure 2.** Network forest plot for progression-free survival (general analysis)

**eFigure 3.** Network forest plot for objective response rate (general analysis)
*Note:* The actual P-score of each regimen should be 1 minus its value displayed in the figure, according to the network calculation in R software.

**eFigure 4.** Network forest plot for hematological adverse events (general analysis)
**eFigure 5.** Network forest plot for non-hematological adverse events (general analysis)

**eFigure 6.** Network forest plot for overall survival (additional analysis)
*Note:* “FC2” was the common comparator.

**eFigure 7.** Network league table for overall survival (additional analysis)
*Note:* Treatments are hierarchically ranked according to their P-score. The higher position in the table a regimen locates at, the better survival benefits it could offer. Values situated at the intersection of specific column and row are the network effect sizes (HR and 95%CI) of row-defining regimen versus column-defining regimen.

**eFigure 8.** Network forest plot for progression-free survival (additional analysis)
*Note:* “FC2” was the common comparator.

**eFigure 9.** Network forest plot for progression-free survival (additional analysis)
*Note:* Treatments are hierarchically ranked according to their P-score. The higher position in the table a regimen locates at, the better survival benefits it could offer. Values situated at the intersection of specific column and row are the network effect sizes (HR and 95%CI) of row-defining regimen versus column-defining regimen.

**eFigure 10.** Network forest plot for objective response rate (additional analysis)
*Note:* The actual P-score of each regimen should be 1 minus its value displayed in the figure, according to the network calculation in R software. “FC2” was the common comparator.

**eFigure 11.** Network forest plot for objective response rate (additional analysis)
*Note:* Treatments are hierarchically ranked according to their P-score. The lower position in the table a regimen locates at, the better response benefits it could offer. Values situated at the intersection of specific column and row are the network effect sizes (RR and 95%CI) of row-defining regimen versus column-defining regimen.

**eFigure 12.** Network forest plot for hematological adverse events (additional analysis)
*Note:* “FC2” was the common comparator.

**eFigure 13.** Network forest plot for hematological adverse events (additional analysis)
Note: Treatments are hierarchically ranked according to their P-score. The higher position in the table a regimen locates at, the better safety benefits it could offer. Values situated at the intersection of specific column and row are the network effect sizes (RR and 95% CI) of row-defining regimen versus column-defining regimen.

eFigure 14. Network forest plot for non-hematological adverse events (additional analysis)
Note: “FC2” was the common comparator.

eFigure 15. Network forest plot for non-hematological adverse events (additional analysis)
Note: Treatments are hierarchically ranked according to their P-score. The higher position in the table a regimen locates at, the better safety benefits it could offer. Values situated at the intersection of specific column and row are the network effect sizes (RR and 95% CI) of row-defining regimen versus column-defining regimen.
Supplementary reference

1. Kim MJ, Kong SY, Nam BH, et al. A randomized phase II study of S-1 versus capecitabine as first-line chemotherapy in elderly metastatic gastric cancer patients with or without poor performance status: clinical and pharmacogenetic results. *Pharmacogenet Genomics* 2018; 28: 23-30. DOI: 10.1097/FPC.0000000000000320.

2. Huang D, Ba Y, Xiong J, et al. A multicentre randomised trial comparing weekly paclitaxel + S-1 with weekly paclitaxel + 5-fluorouracil for patients with advanced gastric cancer. *Eur J Cancer* 2013; 49: 2995-3002. DOI: 10.1016/j.ejca.2013.05.021.

3. Huang DZ, Xiong JP, Xu N, et al. [A phase II prospective randomized controlled trial of weekly paclitaxel combined with S-1 or fluorouracil for advanced gastric carcinoma]. *Zhonghua Zhong Liu Za Zhi* 2012; 34: 865-868. DOI: 10.3760/cma.j.issn.0253-3766.2012.11.015.

4. Cascinu S, Galizia E, Labianca R, et al. Pegylated liposomal doxorubicin, 5-fluorouracil and cisplatin versus mitomycin-C, 5-fluorouracil and cisplatin for advanced gastric cancer: a randomized phase II trial. *Cancer Chemother Pharmacol* 2011; 68: 37-43. DOI: 10.1007/s00280-010-1424-8.

5. Lee JL, Kang YK, Kang HJ, et al. A randomised multicentre phase II trial of capecitabine vs S-1 as first-line treatment in elderly patients with metastatic or recurrent unresectable gastric cancer. *Br J Cancer* 2008; 99: 584-590. DOI: 10.1038/sj.bjc.6604536.

6. Duffour J, Bouche O, Rougier P, et al. Safety of cisplatin combined with continuous 5-FU versus bolus 5-FU and leucovorin, in metastatic gastrointestinal cancer (FFCD 9404 randomised trial). *Anticancer Res* 2006; 26: 3877-3883.

7. Ross P, Nicolson M, Cunningham D, et al. Prospective randomized trial comparing mitomycin, cisplatin, and protracted venous-infusion fluorouracil (PVI 5-FU) With epirubicin, cisplatin, and PVI 5-FU in advanced esophagogastric cancer. *J Clin Oncol* 2002; 20: 1996-2004. DOI: 10.1200/JCO.2002.08.105.

8. Kondo K, Sakamoto J, Nakazato H, et al. A phase III randomized study comparing doxifluridine and 5-fluorouracil as supportive chemotherapy in advanced and recurrent gastric cancer. *Oncol Rep* 2000; 7: 485-490.

9. Wils JA, Klein HO, Wagener DJ, et al. Sequential high-dose methotrexate and fluorouracil combined with doxorubicin--a step ahead in the treatment of advanced gastric cancer: a trial of the European Organization for Research and Treatment of Cancer Gastrointestinal Tract Cooperative Group. *J Clin Oncol* 1991; 9: 827-831. DOI: 10.1200/JCO.1991.9.5.827.

10. Bang YJ, Cho JY, Kim YH, et al. Efficacy of Sequential Ipilimumab Monotherapy versus Best Supportive Care for Unresectable Locally Advanced/Metastatic Gastric or Gastroesophageal Junction Cancer. *Clin Cancer Res* 2017; 23: 5671-5678. DOI: 10.1158/1078-0432.CCR-17-0025.

11. Li W, Zhao X, Wang H, et al. Maintenance treatment of Uracil and Tegafur (UFT) in responders following first-line fluorouracil-based chemotherapy in metastatic gastric cancer: a randomized phase II study. *Oncotarget* 2017; 8: 37826-37834. DOI: 10.18632/oncotarget.13922.

12. Kim JA, Lee J, Han B, et al. Docetaxel/cisplatin followed by FOLFIRI versus the reverse sequence in metastatic gastric cancer. *Cancer Chemother Pharmacol* 2011; 68: 177-184. DOI: 10.1007/s00280-010-1452-4.

13. Gubanski M, Johnsson A, Fernebro E, et al. Randomized phase II study of sequential docetaxel and irinotecan with 5-fluorouracil/folinic acid (leucovorin) in patients with advanced gastric cancer: the GATAc trial. *Gastric Cancer* 2010; 13: 155-161. DOI: 10.1007/s10120-010-0553-4.
14. Bramhall SR, Hallissey MT, Whiting J, et al. Marimastat as maintenance therapy for patients with advanced gastric cancer: a randomised trial. *Br J Cancer* 2002; 86: 1864-1870. DOI: 10.1038/sj.bjc.6600310.