Supplementary methods

Study inclusion criteria

1. Randomized controlled phase II/III trials recruited patients with stage IIIB-IV or previously treated NSCLC who failed first-line therapy;

2. Pathology should be squamous or non-squamous cell NSCLC;

3. Patients were 18 years or older and had an Eastern Cooperative Oncology Group (ECOG) performance status of 0 or 1;

4. Patients did not receive previous immunotherapy or have a history of autoimmune disease;

5. Each treatment arm in the trial should only contained one medication of either immunotherapy or chemotherapy;

Literature search strategy

The following electronic databases were searched to identify all eligible trials for this network meta-analysis: PubMed, Web of Science, Cochrane Library as well as WangFang database and National Knowledge Infrastructure for Chinese literature. Furthermore, it was supplemented by manual search of reference lists of all available primary studies, review articles, meeting reports and relevant books. Searching headlines included non-small cell lung cancer, nivolumab, pembrolizumab, atezolizumab, PD-1 or PD-L1. The search was limited to randomized controlled trials but not English language.

Quality control

To assess the quality of included studies, three investigators (W.Y, M.L and Y.Q.L) independently examined the randomization procedure, estimation of sample size, adoption of blind in study design, allocation concealment, if the intention-to-treat analysis being followed, loss to follow-up and dropout. Jadad/Oxford quality scoring system was adopted to quantify study quality [1]. Any discrepancies would be solved by consensus.

Data extraction

Another three investigators (Y.B.S, D.K.C and Y.Q.L) reviewed the included studies and extracted the data independently. Data on study design, study time, number of patients in each arm, staging information, randomization scheme, follow-up duration, treatment protocol, endpoints and failure patterns were abstracted. Any discrepancies in quality assessment and data extraction were solved by consensus.

Statistical analysis

The primary end-point was OS, defined as the duration from randomization to death from any cause. Second endpoints included progression-free survival (PFS, defined as time from randomization to the first occurrence of disease progression) and objective response (complete response, partial response, stable disease, progressive disease). Survival data were expressed as hazard ratio (HR) and objective response as odds ratio (OR). HR and its variance were directly extracted from the original text. OR and

its variance were calculated through the number of responders (complete response + partial response) and non-responders (stable disease + progression disease) in each treatment arm. Traditional direct meta-analysis was conducted first using Stata 13.0 (StataCorp LP, College Station, TX, USA). We calculated the pooled estimates of HRs or ORs and corresponding 95% confidence intervals (CIs) of direct comparisons between two therapeutic regimens. A two-sided P-value of < 0.05 was considered significant. Heterogeneity across studies was tested by χ^2 test and I^2 statistic along with a forest plot. Statistically significant heterogeneity was defined as a χ^2 P-value < 0.1 or an I^2 statistic $> 50\%$.

The network meta-analysis was planned in the R software (version 3.3.3; R Foundation, Vienna, Austria) using the *netmeta* package [2, 3] and frequentist approach[2]. Logarithmic of HR (logHR) or OR (logOR) and its variance (selogHR or selogOR) would be prepared for statistical data analysis of network meta-analysis. Treatment effects were estimated by HRs or ORs with corresponding 95% confidence intervals (CIs). Heterogeneity or inconsistency between and within designs was established by Q test which was proposed by Rücker et al. [2] to be a generalization of Cochran's test. No heterogeneity existed if $P > 0.1$, and fixed-effects model would be used. In case of significant heterogeneity, the use of random-effects model and the performance of sensitivity analysis would be considered. Forest plots of network meta-analysis were obtained using anti-PD-1, anti-PD-L1 and docetaxel as the reference group, respectively. A P-score, proposed by Rücker and Schwarzer [4] as a frequentist analog to surface under the cumulative ranking curve [5, 6] would be

adopted to rank the treatment arms. P-score would be 100% for the best treatment and 0% for the worst treatment. Overall grade 3-5 toxicities were compared using the χ^2 test and a two-sided *P*-value of < 0.05 was considered significant. Survival and objective response analysis were conducted in intention-to-treat population and toxicity comparison in patients receiving at least one dose of treatment.

Reference

1   Jadad AR, Moore RA, Carroll D, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? Control Clin Trials. 1996; 17: 1-12.

2   Rucker G (2012). Network meta-analysis, electrical networks and graph theory. Res Synth Methods. 2012; 3: 312-24.

3   Rucker G, Schwarzer G, and Krahn U. Netmeta: Network meta-analysis using requentistmethods.http://cran.r-project.org/web/packages/netmeta/index.html.

4   Rucker G, Schwarzer G. Ranking treatments in frequentist network meta-analysis works without resampling methods. BMC Med Res Methodol. 2015; 15:58.

5   Chaimani A, Higgins JP, Mavridis D, et al. Graphical tools for network meta-analysis in STATA. PloS One. 2013; 8: e76654.

6   Salanti G, Ades AE, Ioannidis JP. Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: an overview and tutorial. J Clin Epidemiol. 2011; 64: 163-71.

Table S1. Quality assessment of the included 5 studies using revised Jadad scale.

Study	Randomization process	Allocation concealment	Blinding	Dropout	Jadad score
Nivolumab vs. Docetaxel					
Borghaei et al. 2015	Yes	Yes	No	Yes	5
Brahmer et al. 2015	Yes	Yes	No	Yes	5
Pembrolizumab vs. Docetaxel					
Herbst et al. 2016	Yes	Yes	No	Yes	5
Atezolizumab vs. Docetaxel					
Fehrenbacher et al. 2016	Yes	Yes	No	Yes	5
Rittmeyer et al. 2017	Yes	Yes	No	Yes	5

Figure S1. Flow chart of literature searches and study selection.

PubMed (n = 29)
Web of science (n = 27)
Cochrane Library (n = 24)

48 duplicates excluded

32 titles screened

9 irrelevant studies excluded
5 phase I trials
2 single-arm trials

16 abstracts screened

5 studies having more than one
medications in one arm
2 non-randomized trials
3 trials include other kind of cancer

6 full text assessment

1 study recruited patients with
previously untreated NSCLC

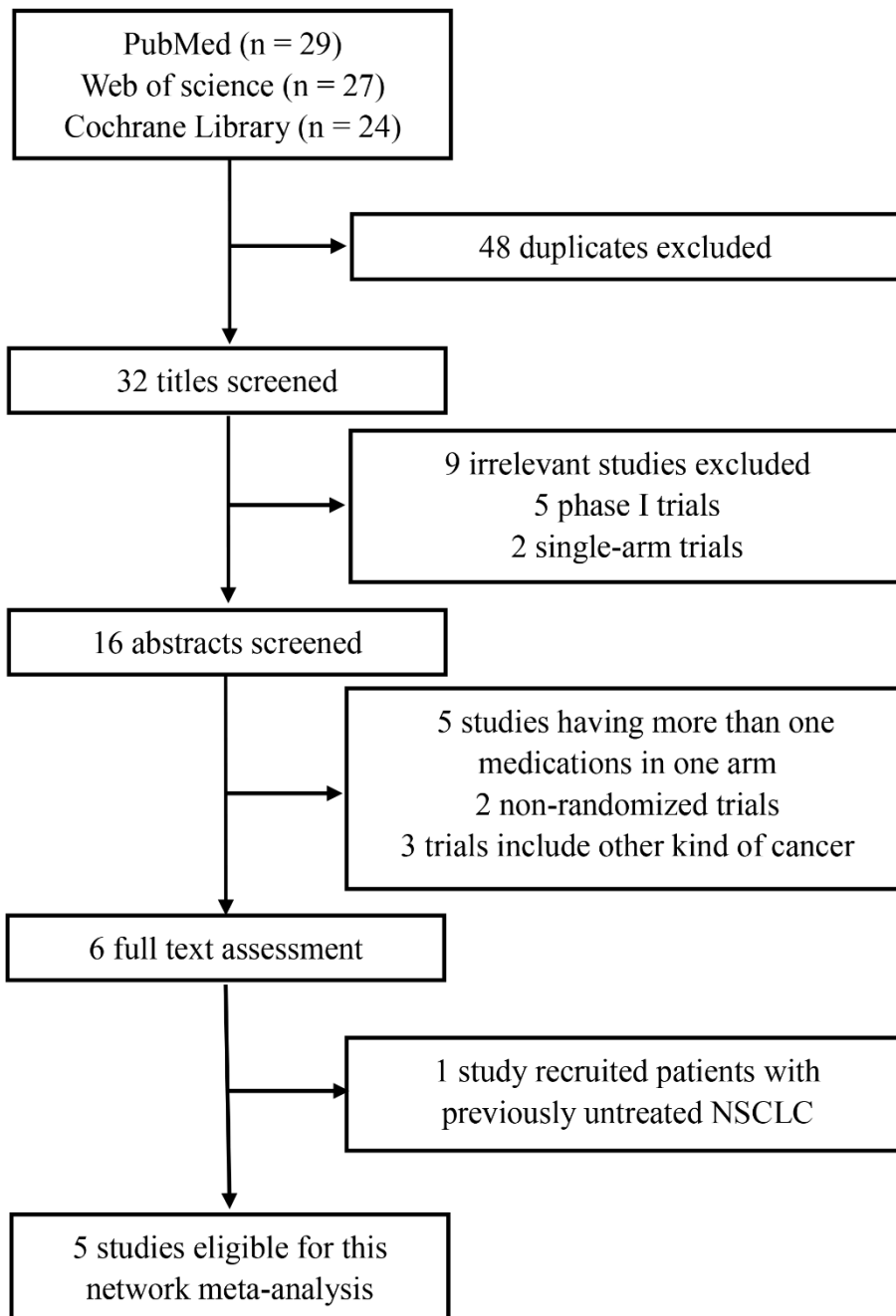5 studies eligible for this
network meta-analysis

Fig. S1. Flow chart of literature searches and study selection.

Figure S2. Graphical presentation of the trial network for overall survival. PD-1 = programmed death 1; PD-L1 = programmed death ligand 1.
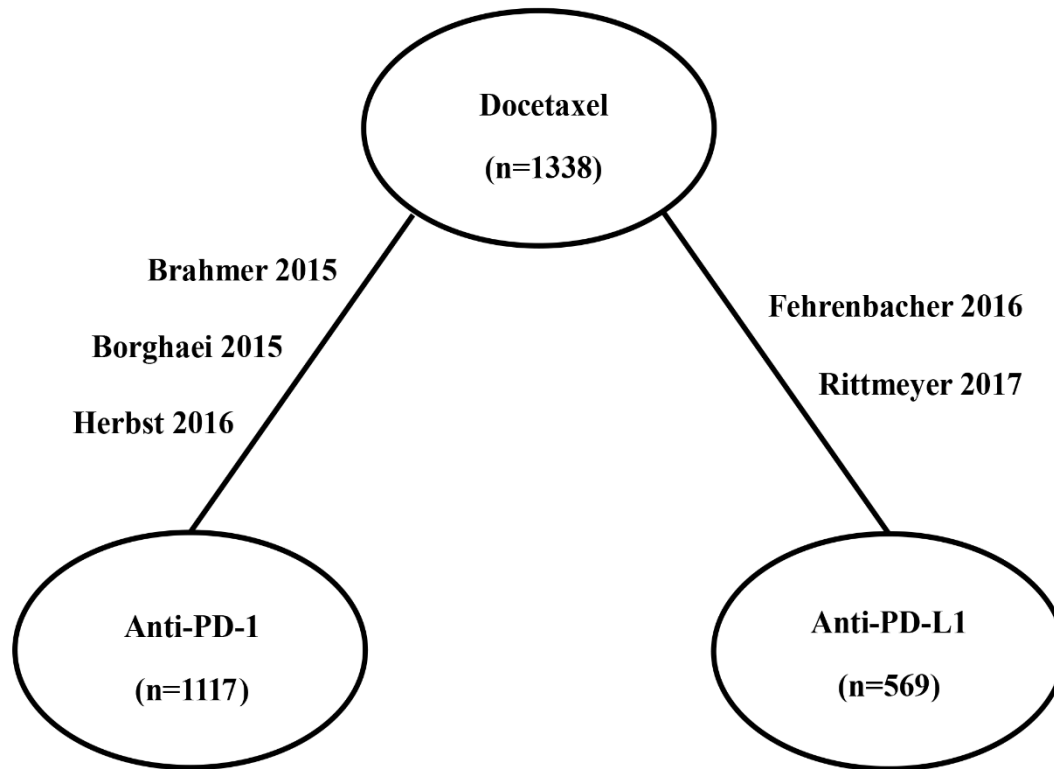

Fig. S2. Graphical presentation of the trial network for overall survival. PD-1 = programmed death 1; PD-L1 = programmed death ligand 1.

Figure S3. Forest plot of network meta-analysis for objective response with different

reference groups. PD-l = programmed death 1; PD-L1 = programmed death ligand 1;

OR = odds ratio; CI = confidence interval.

Treatment	Random Effects Model	OR	95%-CI
Docetaxel		1.00	
Anti-PD-1		0.36	[0.25; 0.52]
Anti-PD-L1		1.02	[0.62; 1.66]
Docetaxel		2.77	[1.94; 3.96]
Anti-PD-1		1.00	
Anti-PD-L1		2.82	[1.54; 5.17]
Docetaxel		0.98	[0.60; 1.60]
Anti-PD-1		0.35	[0.19; 0.65]
Anti-PD-L1		1.00	

0.2   0.5   1   2   5