**Material and Methods**

This meta-analysis followed the Systematic Reviews and Meta-Analyses (PRISMA) [15] .

## Search Strategy

The PubMed, Web of Science, and China National Knowledge Infrastructure databases were searched from inception through December 14, 2018. The keywords used in the search strategy were “(Lymphocytes, Tumor-Infiltrating [MeSH Terms]) AND (melanoma [MeSH Terms]) AND (Prognosis [MeSH Terms] OR Survival Analysis [MeSH Terms] OR mortality [MeSH Terms])”. There were no other limitations in the database search. In addition to the title, abstract, and full text, the reference lists of the identified articles were also reviewed to identify additional potential studies. The availability evaluation and database search were conducted independently by two researchers. Discrepancies between the two researchers were discussed and resolved by consensus.

### *Inclusion and Exclusion Criteria*

The studies included in this meta-analysis met the following criteria: (1) published as original articles; (2) evaluated human subjects; (3) studies of melanoma reporting the prognostic impact of TILs divided into brisk, nonbrisk, and absent; (4) studies providing sufficient data to estimate the hazard ratio (HR) and 95% confidence intervals (CIs); and (5) prognostic value investigated by time-to-event survival analysis with either overall survival (OS), disease-free survival (DFS), or disease-specific survival (DSS).

The exclusion criteria were as follows: (1) TIL classification other than that defined by Clark;(2) studies that did not provide available data to estimate the HRs and 95% CIs; and (3) case reports, letters, reviews, conference reports, abstracts, and animal trials.

### *Data Extraction and Quality Assessment*

The parameters of all eligible studies were collected, including first author surname, publication year, population origin, sample size, follow-up time, clinicopathological characteristics, TIL grade situation, observed outcomes, HRs of TILs for OS, DFS, and DSS, and corresponding 95% CIs. Since the observation end point times of each article were different, we also extracted the HR values up to the 5-year survival time point. When these parameters were not mentioned in the article, we extracted the data from Kaplan-Meier curves by digitizing the curves using the open-source Engauge Digitizer software (<http://digitizer.sourceforge.net/>) and estimating the univariate HR [16]. If univariate and multivariate analyses were both involved in a study, all relevant data were recorded.

The quality of each study was independently assessed by two researchers using the Newcastle-Ottawa Quality Assessment Scale [17]. The Newcastle-Ottawa scale evaluated cohort studies by means of three large blocks of 8 items, including population selection, comparability, exposure evaluation or outcome evaluation. Scores ranged from 0 to 9 for quality assessment, and studies with scores ≥6 were rated as high quality.

### *Statistical Analysis*

HRs and 95% CIs were used to evaluate the risk of death in patients with melanoma at different TIL grades: an observed HR > 1 indicated a worse prognosis in patients with this TIL grade, while an HR < 1 suggested a better prognosis. If the HRs and 95% CIs were reported in the study, we extracted them directly. If the statistical variables were not directly reported in the article, we used the data extracted from the survival curves to calculate logHR and related standard errors（<https://datadryad.org/stash/share/7KpK0H9rMOfVF08dplLwcj9oiiN418OgYKKwneu2yCY>） using a spreadsheet provided by Tierney et al. [16]. The data from Kaplan-Meier survival curves were read by two independent researchers using Engauge Digitizer version 4.1 to reduce variability, and the specific method refers to the description of Zhou et al. [18]. In general, the natural logarithm of HR and its standard error were used in this meta-analysis to make the range of HR symmetric. After logarithmic transformation, an HR of 0 becomes minus infinity, an HR of 1 becomes 0, and an HR of infinity remains infinity [19].

Statistical heterogeneity was assessed by visual inspection of forest plots, by χ2 tests to assess the *p* value, and by *I*2 statistics [20, 21]. Statistical methods for meta-analysis included the fixed-effect model and the random-effect model. The fixed-effect model refers to the calculation model of combined effect quantity in the meta-analysis, which assumes that all the observed variations between studies are caused by chance. These studies assume to measure the same overall effect. A random-effect model is a model that estimates the uncertainty (CI) of the results through the in-study sampling error (variance) and interstudy variation in the statistical meta-analysis. When the included studies have heterogeneity except accidental chance, the random-effect model will give a wider CI than the fixed-effect model [22]. The analysis model was selected according to the heterogeneity. *p* values <0.10 and/or *I*2 >50% indicated the presence of heterogeneity, and a random-effect model (DerSimonian-Laird method) was applied to partially eliminate the effects of heterogeneity. Otherwise, the fixed-effect model (Mantel-Haenszel method) was used.

The number of studies included in the systematic review will have an impact on the effectiveness of the funnel plot in detecting bias. If too few studies are included, the detection efficiency of the funnel plot method decreases correspondingly. Funnel plots should be used for reviews with sufficient numbers of included studies. Unfortunately, since fewer than 10 articles were included in each of our analyses, we did not evaluate publication bias. All analyses were performed using Review Manager version 5.3 (The Cochrane Collaboration, Copenhagen: The Nordic Cochrane Centre, 2012) and STATA version 12.0 (STATA Corporation, College Station, TX, USA), with significance defined as *p* values <0.05.