The Prognostic Value of Circulating Tumor Cells in Patients with Melanoma: A Systematic Review and Meta-analysis

Simone Mocellin,1,2 Dave Hoon,3 Alessandro Ambrosi,4 Donato Nitti,1 and Carlo Riccardo Rossi1

Abstract

Background: The detection of circulating tumor cells (CTC) in patients with melanoma represents an appealing prognostic tool, but no consensus exists on this topic. We aimed to comprehensively and quantitatively summarize the evidence for the use of CTC to predict patients’ clinical outcome.

Methods: Fifty-three studies enrolling 5,433 patients were reviewed. Correlation of CTC status with tumor-node-metastasis disease stage and patients’ overall (OS) and progression-free (PFS) survival was assessed by means of association statistics and meta-analysis, respectively.

Results: CTC status correlated with both tumor-node-metastasis stage (stage I, 32%; stage II, 41.7%; stage III, 41.1%; stage IV, 47.4%; \( P_{\text{trend}} < 0.0001 \)) and survival (OS: hazard ratio, 2.42; 95% confidence interval, 1.7-3.45, \( P < 0.0001 \); PFS: hazard ratio, 2.45; 95% confidence interval, 1.78-3.38; \( P < 0.0001 \)). However, statistical heterogeneity was significant for both OS and PFS, likely underscoring the wide variability in study design. Furthermore, CTC positivity rates in early stages were higher and in the metastatic setting were lower than expected, which indicates an unsatisfactory accuracy of currently available CTC detection assays.

Conclusions: Our findings suggest that CTC might have a clinically valuable prognostic power in patients with melanoma. However, the heterogeneity of the studies thus far published warrants caution not to overestimate the favorable results of pooled data.

The search for minimal residual disease in the peripheral blood is routinely done for the therapeutic management of patients with hematologic malignancies, as the bloodstream is the “physiologic” milieu for this kind of tumors. By contrast, the biological significance of circulating tumor cells (CTC) in solid tumors is still debated (1). The largest experience in this field has been gained in patients affected with melanoma, the first solid tumor recognized to shed CTC by PCR-based amplification of the melanocyte-specific tyrosinase gene (2). Although tens of studies have been done and thousands of patients have been enrolled, there is no consensus on CTC biological significance in patients with melanoma (as well as other solid malignancies) and thus on the use of CTC in the routine clinical setting. At this point, a question may arise whether—based on the available findings—it is justified to further explore the prognostic power of CTC.

With the aim of gaining a better insight into the prognostic value of CTC in patients with melanoma, we undertook a systematic review of published studies and used standard meta-analysis techniques to pool together and quantitatively summarize the available evidence. We intended to assess the quality of published series, to identify technical factors influencing CTC detection rates, and to evaluate the effect of CTC status (presence versus absence) on both indirect [tumor-node-metastasis (TNM) stage] and direct [overall (OS) and progression-free (PFS) survival] indicators of patients’ clinical outcome.

Materials and Methods

Search strategy, eligibility criteria, and data extraction. A systematic review of original articles analyzing the prognostic value of CTC in melanoma patients was done by searching Medline, Embase, Cancerlit, and Cochrane databases. The search strategy included the following keywords variably combined: “melanoma,” “circulating tumor (melanoma) cells,” “disseminated (melanoma) tumor cells,” “minimal residual disease,” “prognosis,” “prognostic marker,” “peripheral blood,” “polymerase chain reaction,” and “immunomagnetic cell enrichment.” Original and review articles published until November 2005 were sought, considering the latter as an additional source of original works otherwise overlooked. When appropriate, cited references from selected articles were also reviewed. Only studies published in peer-reviewed journals were included; data from letters and meetings abstracts were not considered eligible. Authors were contacted whenever data not reported were useful to include the study into this systematic review or to rule out data published in different papers but regarding overlapping series.
Studies were eligible if disease stage and/or survival (OS and/or PFS) data were reported in melanoma patients stratified by CTC status (presence/positive versus absence/negative). The major restriction criterion was the number of patients enrolled, which had to be ≥30. Articles describing the detection of free RNA or DNA in plasma samples were not considered. Care was taken to account for overlapping and duplicate data sets. To this aim, we identified articles that included the same cohort of patients by reviewing between-study similarity, investigators, recruitment period, and inclusion criteria. When the same authors reported results obtained from the same series of patients in multiple publications, only the largest or the most informative (e.g., reporting survival data) were included in the analysis. Duplicate reports were included in the specific analyses only if they applied different methods, tumor markers, or disease stage, and if they added survival analysis previously not done. A cohort of patients was not included more than once in the same analysis.

A comprehensive database was designed to ensure that all data needed for analysis were publicly available (Supplementary Table S1). Data were extracted by two investigators (S.M. and C.R.R.) to ensure homogeneity of data collection and to rule out the effect of subjectivity in data gathering and entry. Disagreements were resolved by iteration, discussion, and consensus. To unravel potential systematic biases, a third investigator (D.N.) did a concordance study by independently reviewing 50% of the eligible studies randomly chosen: Complete concordance was reached for all variables assessed. All statistical analyses were done with the supervision of a biostatistician (A.A.).

**Statistical analysis.** Three main methods were used to summarize results. (a) weighted multiple regression was used to evaluate the influence of several factors (publication year, number of TNM stages considered in the study, enrollment of stage IV patients, in vitro sensitivity assay, blood sample volume, number of blood samples per patient, RNA extraction method, PCR technique, and number of tumor markers used to detect CTC) on CTC detection rates; the square root of the sample size was used as weighting factor, and the arcsin transformation of the CTC positivity rate of each study was considered the dependent variable; (b) to assess the association between CTC status and TNM stages, the Cochran-Armitage trend test was fitted to data pooled across studies; (c) standard meta-analysis methods (3, 4) were applied to evaluate the overall effect of CTC status on patients’ survival.

The association between CTC and survival (OS and PFS) was derived as a weighted average of study-specific estimates of the hazard ratio (HR, observed – expected cases) / (observed – expected cases), using the inverse of variance as the weighting factor (3, 4). The natural logarithm of HR (lnHR) and the corresponding SE were used as data points for the meta-analysis. In studies evaluating the independence of CTC prognostic value by means of multivariate survival analysis (Cox proportional hazards regression model), HRs and confidence intervals (CI) are usually reported. For studies performing only univariate survival analysis (log-rank test–based comparison of Kaplan-Meier survival curves), HRs and 95% CIs were calculated from survival curves adopting a hierarchical series of steps as per Parmar et al. (5). Briefly, within a given study, survival probability was calculated for each group (CTC-positive and CTC-negative patients) were calculated for a minimum of six nonoverlapping time points. Then, the effective number of patients alive, at risk, and censored was calculated for each time point. The overall lnHR for each study was considered a weighted sum of the individual estimates of the lnHR during each time interval \( t \), where the weights are inversely proportional to the variance of each estimate \( \sigma^2 \), as described by the following expression:

\[
\text{lnHR} = \frac{\sum \ln(\text{HR}(t)) / \sum 1 / \sigma^2}{\sum 1 / \sigma^2}
\]

The corresponding SE was calculated according to the following formula: \( 1 / \sqrt{V} \), where \( V \) is the variance of the log-rank statistic.

Because a mixture of log-rank and Cox model estimates were obtained from the eligible studies, results were combined using the generic inverse variance method (3, 4). Meta-analysis was first done using the fixed-effects model, which assumes that all the studies share the same common prognostic effect. Results (effect sizes) consistency among studies was investigated by means of two heterogeneity tests: the \( \chi^2 \)-based Cochran’s Q test and the I² statistic \( (Q - df) / Q \times 100 \), where Q is the Cochran statistic and \( df \) are the degrees of freedom, that is the number of studies minus 1; it indicates the percentage of the variability in effect estimates due to heterogeneity rather than sampling error. To be more conservative, we considered that heterogeneity was present when the Cochran’s Q test \( P \) value was <0.1. In addition, inconsistency across studies was quantified by means of I² statistic, which is generally considered significant (i.e., heterogeneity has a significant effect on meta-analysis) for values ≥50%. In case of heterogeneity, meta-analysis was done by applying the random-effects model (6), which assumes that studies were a random sample of a hypothetical population of studies and assigns a weight to each study taking into account variance within and between studies.

The extent to which the combined risk estimate might be affected by individual studies was assessed by consecutively omitting every study from the meta-analysis. Subgroup analysis considering more homogeneous sets of studies was adopted as an additional sensitivity test. Furthermore, to identify potential sources of heterogeneity, random-effects meta-regression was also implemented, as per Thompson and Sharp (7). In brief, potential explanatory (independent) variables (covariates, i.e., study features that might affect the magnitude of the effect estimate) were regressed against the outcome (dependent) variable (lnHR), after that each study was weighted by the precision of its respective effect estimate (1 / SE[lnHR]). Residual heterogeneity (not explained by the covariates) was incorporated in the weight formula (1 / [SE + \( t^2 \)]) by allowing an additive between-study variance component \( t^2 \), which was estimated according to the following formula:

\[
\tau^2 = \frac{Q - (k - 2)}{I^2} \times \frac{1}{P(w, x)}
\]

where \( Q \) is the heterogeneity statistic, \( k \) is the number of studies, \( w \) is the estimate precision, \( x \) is the covariate, and

\[
P(w, x) = \sum w - \sum w^2 \Sigma wx^2 - \Sigma w^2 x \Sigma wx + \sum w^2 x^2 - \Sigma wx^2 - (\Sigma wx)^2\]

Publication bias (linked to the fact that negative trials are cited less frequently and are therefore more likely to be missed in the search for relevant studies) was sought with the “funnel plot” technique (based on a graph plotting effect estimates against sample size; ref. 8). Funnel plot asymmetry on the natural logarithm scale of the HR was formally investigated with Egger’s method, a linear regression approach in which the standard normal deviate (the HR divided by its SE) is regressed against the estimate precision \( 1 / SE[\text{lnHR}] \). The effect of publication bias on the pooled effect was assessed by the “trim-and-fill” method described by Duval and Tweedie (10), using both \( L_0 \) and \( R_0 \) as estimators of the potentially missing studies; to this aim the following formulas were used:

\[
L_0 = \frac{4 \bar{t}_n - n(n + 1)}{2n - 1} \]

and

\[
R_0 = \gamma_{.05} - L_0 \]

where \( n \) is the number of published studies and \( \bar{t}_n \) is the “trimmed” rank test statistic.

Study quality was quantitatively assessed by comparing the number of potential sources of heterogeneity, meta-analysis was done by applying the random-effects model (6), which assumes that studies were a random sample of a hypothetical population of studies and assigns a weight to each study taking into account variance within and between studies.

The extent to which the combined risk estimate might be affected by individual studies was assessed by consecutively omitting every study from the meta-analysis. Subgroup analysis considering more homogeneous sets of studies was adopted as an additional sensitivity test. Furthermore, to identify potential sources of heterogeneity, random-effects meta-regression was also implemented, as per Thompson and Sharp (7). In brief, potential explanatory (independent) variables (covariates, i.e., study features that might affect the magnitude of the effect estimate) were regressed against the outcome (dependent) variable (lnHR), after that each study was weighted by the precision of its respective effect estimate (1 / SE[lnHR]). Residual heterogeneity (not explained by the covariates) was incorporated in the weight formula (1 / [SE + \( t^2 \)]) by allowing an additive between-study variance component \( t^2 \), which was estimated according to the following formula:

\[
\tau^2 = \frac{Q - (k - 2)}{I^2} \times \frac{1}{P(w, x)}
\]

where \( Q \) is the heterogeneity statistic, \( k \) is the number of studies, \( w \) is the estimate precision, \( x \) is the covariate, and

\[
P(w, x) = \sum w - \sum w^2 \Sigma wx^2 - \Sigma w^2 x \Sigma wx + \sum w^2 x^2 - \Sigma wx^2 - (\Sigma wx)^2\]

Publication bias (linked to the fact that negative trials are cited less frequently and are therefore more likely to be missed in the search for relevant studies) was sought with the “funnel plot” technique (based on a graph plotting effect estimates against sample size; ref. 8). Funnel plot asymmetry on the natural logarithm scale of the HR was formally investigated with Egger’s method, a linear regression approach in which the standard normal deviate (the HR divided by its SE) is regressed against the estimate precision \( 1 / SE[\text{lnHR}] \). The effect of publication bias on the pooled effect was assessed by the “trim-and-fill” method described by Duval and Tweedie (10), using both \( L_0 \) and \( R_0 \) as estimators of the potentially missing studies; to this aim the following formulas were used:

\[
L_0 = \frac{4 \bar{t}_n - n(n + 1)}{2n - 1} \]

and

\[
R_0 = \gamma_{.05} - L_0 \]

where \( n \) is the number of published studies and \( \bar{t}_n \) is the “trimmed” rank test statistic.
<5%. When considering the association between CTC status and TNM stage, the expected sample sizes were estimated under the following assumptions: (a) equal sample size of stage populations and (b) expected CTC positivity rates equivalent to the 5-year mortality rate per stage; (c) HR, 2; and (d) median PFS, 50% of median OS.

Meta-analysis was conducted using the RevMan software version 4.2 (The Cochrane Collaboration, Oxford, United Kingdom). Other statistical analyses were done in the “R” environment (release version 2.2.0) and using the SPSS statistical package (SPSS, Inc., Chicago, IL, release version 13.0). Values of 95% CI were used for all analyses. Probability values <5% were considered significant.

Results

Characteristics of identified studies

Literature search with the above-mentioned keywords yielded 209 articles; of these, 123 articles reporting on CTC detection in melanoma patients were reviewed in detail (Fig. 1). Excluding reviews and those reporting only on methodologic aspects, 53 original articles met all inclusion criteria and represent the source of the data analyzed here (12–64). A list of excluded articles is reported in Supplementary Table S2.

The included 53 studies encompassed 5,433 patients affected with TNM stage I to IV cutaneous ($n = 5,388; 99.2\%$) and uveal ($n = 45; 0.8\%$) melanoma. The mean number of patients enrolled was 102 (range 20–299), with 20 (37.7\%) and 6 (11.3\%) studies enrolling ≥100 and 200 patients, respectively. Considering the studies that reported patient demographics (36 of 53, 67.9\%), there were 1,823 females (49.6\%) and 1,854 males (50.4\%), and the mean age was 51.6 years (range 36–57 years).

In 47 studies (88.7\%), 1,361 controls (healthy persons or patients with cancers other than melanoma) were enrolled (mean 29, range 1–89).

All TNM stages (I–IV) were considered in 31 studies (58.5\%); stage I, II, and III were considered in seven (13.2\%), eight (15.1\%), and seven (13.2\%) studies, respectively. Overall, stage IV patients were enrolled in most series (41 of 53, 77.3\%).

Spiking experiments to test in vitro the sensitivity of CTC detection methods were reported in 43 of 53 studies (81.1\%). In 30 studies (69.8\%), it was possible to find one melanoma cell in $1 \times 10^6$ peripheral blood mononuclear cells or in 5 mL whole blood; in the other 13, the sensitivity was lower.

The mean blood sample volume was 10.7 mL (range 1.5–50 mL), with 35 of 51 evaluable studies using ≥10 mL (68.6\%). Single and multiple blood withdrawals per patient were done in 28 (52.8\%) and 25 (47.2\%) studies, respectively. Considering 52 evaluable studies, CTC were searched in peripheral blood samples after density gradient cell separation and RBC lysis in 26 (50\%) and 7 cases (13.4\%), respectively, whereas whole blood was used in 17 (32.7\%); a miscellany of these approaches was adopted in two studies (3.9\%). CTC detection was based on PCR in all but one study, in which the authors used immunomagnetic cell enrichment combined with light microscopy (56). Among PCR-based methods, nested PCR was the most widely used (42 of 52, 80.7\%). Single-round PCR was used in six cases (11.5\%), quantitative real-time PCR in three (5.8\%), and a mix of single and nested PCR in one (2\%). PCR was enhanced with Southern blot, ELISA, and electrochemoluminescence analysis of amplicons in nine (17.3\%), two (3.8\%), and one (1.9\%) study, respectively.

CTC identification hinged upon single and multiple (mean 2.9, range 2–5) tumor markers in 30 (56.6\%) and 23 (43.4\%) cases, respectively. In the only study using the cytometric approach (which hinges upon visualization of intact CTC, as opposed to PCR-based methods where normal peripheral blood mononuclear cell and potentially contaminating CTC are lysed to extract the genetic material and assess the expression of tumor-related genes), immunomagnetic cell enrichment was based on a monoclonal antibody directed to melanoma-associated chondroitin sulfate proteoglycan, and light microscopy was used to recognize circulating melanoma cells by morphology only. In the other studies, PCR was directed to amplify genes coding for nine different melanoma-associated markers: tyrosinase (49 of 52 studies; 94.2\%), melanoma antigen recognized by T cells-1 (MART1/Melan-A; 19 of 52, 36.5\%), melanoma antigen gene family (7 of 52, 13.5\%), glycoprotein gp100/pmml7 (5 of 52, 9.6\%), melanotransferrin (p97; 3/52, 5.8\%), tyrosinase-related proteins (3 of 52, 5.8\%), β-N-acetyl-galactosaminy1-transferase (3 of 52, 5.8\%), melanoma cell adhesion molecule MUC18 (2 of 52, 3.8\%), paired box homeotic gene transcription factor-3 (2 of 52; 3.8\%), and melanoma inhibitory activity protein (1 of 52, 1.9\%).

Effect of study design variables on CTC detection

Data regarding CTC positivity rates were extracted according to the following rules: When the expression of more than one tumor marker per sample was assessed and/or more than one blood sample per patient was withdrawn, we considered positive a case positive to at least one marker in at least one sample. If these data were unavailable, either the marker with the best specificity (lowest false-positive rates among controls) or the one correlating with stage or survival was chosen. Similarly, the sample series (e.g., before versus after treatment) correlating with stage/survival was chosen. Whenever possible, we evaluated per patient (not per sample) results.
Considering that in five studies results were reported (partially or systematically) on a per sample basis (13, 24, 26, 30, 54), CTC were detected in 2,231 of 5,644 cases (39.5%) and in 8 of 1,361 controls (false-positive rate, 0.6%). The latter value might be underestimated, as controls were sampled once in virtually all studies, whereas patients were sampled multiple times in 25 of 53 series.

To investigate whether study design or technical features might influence CTC detection, nine independent variables were analyzed: publication year, in vitro sensitivity (one versus more melanoma cells per 1 × 10⁶ peripheral blood mononuclear cells or per 5 ml peripheral blood), number of TNM stages considered (one to four), inclusion of stage IV patients (yes versus no), sample volume (milliliters), number of samples (single versus multiple), number of tumor markers (single versus multiple), type of PCR (nested versus other), and type of blood manipulation (density gradient cell separation versus other). Considering all 53 studies, after stepwise model variable selection and case-wise deletion of missing values, only two variables significantly and independently affected CTC positivity rates, the use of multiple markers being correlated with higher CTC detection rates (odds ratio, 2.14; 95% CI, 1.1-4.16; P: 0.025).

**CTC prognostic value**

*Correlation with disease stage.* Among the 5,433 enrolled patients, 9 will not be considered in the following TNM stage-related analysis because they were in stage 0 (*in situ* melanoma; ref. 49): these patients, however, will be considered in the following paragraph regarding survival analysis, as the authors included them in the Cox model. In older publications where a three-stage classification was adopted, stage I, II, and III have been considered to correspond to currently accepted stage I-II, III, and IV, respectively. Therefore, of 5,424 patients, 984 were in stage I (18.1%), 329 in stage I-II (6.0%), 1,022 in stage II (18.8%), 356 in stage II-III (6.6%), 1,376 in stage III (25.4%), 126 in stage III-IV (2.3%), 46 in stage II-IV (0.8%), and 1,185 in stage IV (21.8%).

Taking into consideration 46 studies (*n* = 4,974) reporting on two or more TNM stages, the number of patients enrolled was lower than that expected for trend test analysis only in seven studies (refs. 28, 29, 33, 45, 55, 61, 64; Fig. 2; Supplementary Fig. S1). The correlation between CTC detection and TNM stage was evaluated by combining the data of each single stage across series, which was possible in 46 of 53 studies (86.8%). Taking into account that CTC positivity was expressed on a per sample basis (or as a mixture of patients and samples) in five series (13, 24, 26, 30, 54), CTC positivity rates were 32% (319 of 995; 95% CI, 29.1-35.0%) in stage I, 41.7% (413 of 989; 95% CI, 38.6-44.9%) in stage II, 41.1% (582/1,414; 95% CI, 38.5-43.7%) in stage III, and 47.4% (597 of 1,258; 95% CI, 44.6-50.2%) in stage IV. The Cochran-Armitage test revealed a significant linear trend for CTC positivity among the four stages (*P* < 0.0001), although the association seems to be linked mostly to the difference between the extreme stages (I and IV). However, CTC detection rates were remarkably similar across TNM stages (range 32-47.4%) and significantly different from those one would expect based on the survival rates observed in patients with different stages of disease (Fig. 3; ref. 11).

*Correlation with survival.* The studies correlating CTC status with death or relapse rates by means of χ² statistics (13, 23, 27, 28, 31, 39, 60) were not considered in the following analysis, but the results are reported in Supplementary Table S1.

Survival analysis according to CTC status was done in 22 of 53 studies (41.5%) encompassing 2,401 patients (2,401 of 5,433: 44.2%; refs. 14, 15, 22, 25, 29, 34, 36, 37, 41, 45, 46, 48-50, 53, 55-58, 62-64). The mean follow-up was 35.4 months (range 9-75 months). Patient distribution by stage was as follows: stage 0: 9 (0.4%); stage I: 403 (16.8%); stage I-II: 16 (0.7%); stage II: 528 (22%); stage II-III: 356 (14.8%); stage III: 628 (26.1%); stage II-IV: 46 (1.9%); stage IV: 415 (17.3%). OS and PFS survival was analyzed in 13 (1,078 patients; refs. 14, 22, 29, 36, 41, 45, 53, 56-58, 63, 64) studies, respectively: in seven series (29, 36, 45, 53, 57, 58, 64), the authors analyzed both OS and PFS.

As regards OS, no study reported the estimated sample size and 3 of 13 (23%) reached the required sample size calculated under the assumptions described above (Fig. 2; Supplementary Fig. S2).
Univariate analysis alone (log-rank based comparison of Kaplan-Meier curves) was done in 4 of 13 studies (30.8%; refs. 14, 29, 37, 41). Multivariate analysis (Cox proportional hazards model) was done in the remaining nine series (69.2%; refs. 22, 36, 45, 53, 56–58, 62, 64). In one study (45), the authors stated that CTC status was not an independent predictor of OS, but the statistic necessary for meta-analysis (HR, CI) was not reported; therefore, in this case, we extracted the survival data from the Kaplan-Meier curve.

Singularly taken, 11 of 13 series (84.6%) reported a significant correlation between OS and CTC status (three at univariate and eight at multivariate analysis, respectively). In contrast, two studies (15.4%; refs. 41, 45) showed negative results (at univariate and multivariate analysis, respectively). Of note, these negative results come from studies underpowered for type-\(\beta\) error (power \(\leq 0.3\)).

Meta-analysis (Fig. 4) of pooled data showed a significantly increased risk of death in patients with CTC positivity (HR, 2.42; 95% CI, 1.7–3.45; \(P < 0.0001\)). Because heterogeneity tests were positive (Q statistic \(P = 0.001\); \(I^2 = 63.5\%\)), the random-effects model was applied to estimate the overall effect. The results of subgroup analysis, including eight studies reporting the statistic of the multivariate Cox model, were similar to these findings (HR, 3.02; 95% CI, 1.92–4.75; \(P < 0.0001\)), suggesting that the type of statistic (univariate versus multivariate) had no remarkable effect on the outcome of the analysis. Moreover, the "leave-one-out" procedure revealed that no single study accounted for heterogeneity.

While performing meta-regression, we considered the following covariates as potential sources of heterogeneity: year of publication, sample size, number of stages (one to four), stage IV inclusion, type of PCR (nested versus other), type of blood manipulation (density gradient cell separation versus other), blood sample volume (milliliters), number of blood samples per patient (single versus multiple), number of tumor markers used (single versus multiple), and CTC positivity rate. Meta-regression did not identify any covariate significantly associated with HR estimates; however, because this analysis was characterized by a low power, we cannot rule out that with larger sample sizes one or more of the covariates we considered might significantly affect the meta-analysis outcome.

The weights of the studies (mean 7.7%, SE 0.66) were not statistically different from the expected average weight (one-sample \(t\) test \(P = 0.99\)). The lack of a dominant study leading the results of meta-analysis was confirmed by the leave-one-out procedure, which generated HR estimates (mean HR, 2.45; range 2.19–2.62) very close to those obtained with all 13 studies.

The Egger’s test (intercept \(-0.34; P = 0.002\) showed funnel plot asymmetry, which suggested publication bias. Applying the trim-and-fill procedure, three studies were estimated to be missing, and the adjusted estimate of overall HR resulted 1.74 (95% CI, 1.44–2.11, \(P < 0.0001\)).

Considering PFS, no study reported the estimated sample size and seven (46.7%; refs. 25, 36, 46, 48–50, 57) reached the required sample size calculated on the basis of the conditions made for the above described assumption (Fig. 2; Supplementary Fig. S2).

Univariate analysis alone was done in 2 of 15 studies (13.3%; refs. 29, 46). Multivariate analysis based on the Cox model was done in the remaining 13 cases (86.7%; refs. 25, 34, 36, 45, 48–50, 53, 55, 57, 58, 63, 64).
In one study (15), the authors stated that CTC status was an independent predictor of OS; however, they did not report the statistic necessary for meta-analysis (HR, CI). Therefore, we did not include this series in the following meta-analysis also in the light of the fact that a partially overlapping series published later by the same group (29) is included in the following analysis.

Singularly taken, 11 of 15 series (73.3%) reported a significant correlation between PFS and CTC status; in contrast, four studies (26.7%; refs. 46, 49, 57, 63) showed negative results (three at multivariate and one at univariate analysis; ref. 46). For one of them, the study power was \( \approx 0.5 \) (63). In another study enrolling stage 0 to IV patients (49), on the exclusion of stage IV cases, the multivariate analysis showed that CTC detection was independently associated with a worse prognosis.

Our meta-analysis (Fig. 5) confirmed a significantly increased risk of disease progression in patients with CTC positivity (HR, 2.45; 95% CI, 1.78-3.38; \( P < 0.0001 \)). Because heterogeneity was significant (Q statistic \( P = 0.0001; I^2 = 66.7\% \)), the random-effects model was applied to estimate the overall effect. The results of subgroup analysis, including 13 studies reporting the statistic of the multivariate Cox model, were similar to these findings (HR, 2.56; 95% CI, 1.96-3.34; \( P < 0.0001 \)), suggesting that the type of survival statistic had no remarkable effect on the outcome variable. According to the leave-one-out procedure, no single study accounted for heterogeneity.

Considering the same covariates analyzed for OS, metaregression did not identify potential sources of heterogeneity. The weights of the studies (mean 6.6%, SE 0.56) were not statistically different from the expected average weight (one-sample \( t \) test \( P = 0.94 \)). The lack of a dominant study was confirmed by the leave-one-out procedure, which generated HR estimates very close to those obtained with all 15 studies (mean HR 2.43; range 2.33-2.62).

Egger’s test suggested that publication bias was present (intercept -1.29; \( P < 0.0001 \)). The trim-and-fill analysis revealed that one study might be missing and that if this were published, the adjusted HR would be 2.34 (95% CI, 1.71-3.21; \( P < 0.0001 \)).

**Discussion**

Current prognostic systems (e.g., TNM staging system, primary tumor molecular features, and serum biomarkers) are largely inadequate for an optimal therapeutic management of melanoma patients, as shown by the fact that many patients radically treated (i.e., without evidence of residual disease according to conventional methods) do experience disease recurrence, whereas others currently considered eligible for, and treated with, adjuvant therapy do not require it, as they will not recur (e.g., \( \approx 50\% \) of patients with TNM stage III melanoma; ref. 11). Detection of CTC might represent an ideal prognostic tool as the presence of malignant cells in the peripheral blood is a necessary (although not sufficient) phenomenon for the development of metastatic tumor spread. Unfortunately, there is no consensus on whether CTC detection has a prognostic power reliable enough to be implemented in the routine clinical practice.

![Fig. 5. Forest plot of HRs for PFS from 15 studies (1,871 patients). M, multivariate analysis (Cox proportional hazard model); U, univariate analysis (log-rank test).](www.aacrjournals.org)

- **Authors**
  - Curry et al. (25) M 1.0900 0.3520
  - Mellado et al. (29) U 1.3860 0.5600
  - Hoon et al. (34) M 1.2350 0.5400
  - Proebstle et al. (36) M 1.0850 0.3500
  - Gogas et al. (45) M 1.2996 0.6110
  - Garbe et al. (48) U -0.3000 0.2340
  - Osoka-Arita et al. (48) M 1.4600 0.3660
  - Patrini et al. (49) M 0.3290 0.2680
  - Reynolds et al. (50) M 0.9200 0.3700
  - Wascher et al. (53) M 1.1370 0.4670
  - Mocellin et al. (55) M 0.7227 0.2474
  - Domingo-Domenech et al. (57) M 0.3570 0.1929
  - Koyanagi et al. (58) M 2.3220 0.8500
  - Schmidt et al. (63) M 1.1130 0.6100
  - Vott et al. (64) M 1.3800 0.3760

- **Log HR**
  - 2.94 [1.48, 5.87]
  - 4.00 [1.33, 11.98]
  - 3.44 [1.19, 9.91]
  - 2.96 [1.49, 5.88]
  - 3.67 [1.11, 12.12]
  - 0.74 [0.47, 1.17]
  - 4.31 [2.1, 8.62]
  - 1.39 [0.82, 2.35]
  - 2.51 [1.22, 5.18]
  - 3.12 [1.25, 7.76]
  - 2.06 [1.27, 3.35]
  - 1.43 [0.98, 2.09]
  - 10.20 [1.83, 53.95]
  - 3.10 [0.94, 10.24]
  - 3.87 [1.91, 8.31]

- **HR**
  - 2.45 [1.78, 3.38]

- **CI**
  - 100.00

- **Total (95% CI)**

- **Test for heterogeneity:** \( \chi^2 = 42.06, df = 14, P = 0.0001, I^2 = 66.7\% \)
- **Test for overall effect:** \( Z = 5.47 (P < 0.0001) \)
The present analysis is based on a large pool of clinical series (n = 5,433 patients) and substantially differs from the other two meta-analyses thus far published (65, 66), which considered smaller series (1,799 and 861 patients, respectively), evaluated studies analyzing only one melanoma marker (tyrosinase), and addressed the correlation of CTC status with TNM stage, but not with survival.

Overall, our results support the hypothesis that CTC could play a clinically useful prognostic role in patients with melanoma. In fact, most studies reported a significant correlation between CTC detection and patients’ survival, and most of them showed CTC-independent prognostic value at multivariate analysis. In addition, meta-analysis of pooled data confirmed that CTC represent a significant meta-risk for both OS and PFS, even after adjustment for publication bias. Nevertheless, our findings also hint some considerations that suggest caution in interpreting this favorable outcome and, on the other hand, might pave the way to the design of more informative studies.

The significant heterogeneity found in the meta-analysis for both OS and PFS is likely to be sustained by the number of differences existing between studies, such as stage of disease (from one to four), timing of blood withdrawal (before versus during versus after treatment; with versus without evidence of disease), methods of CTC detection (PCR-based versus cytometric), technical features (density gradient cell separation versus RBC lysis versus whole blood; with versus without cell enrichment; different PCR primers for the same marker; different PCR and cytometry methods), type and number of tumor markers analyzed, definition of risk (CTC presence/absence versus CTC cutoff values; refs. 34, 56, 58), clinical end point (correlation with disease stage or survival), and statistical analysis (univariate versus multivariate survival analysis, different covariates investigated at multivariate analysis). Moreover, intrastudy variability (e.g., enrollment of patients treated with different regimens or blood samples withdrawn at different time points of the patients’ treatment schedule) made it virtually impossible to assess the effect of one of the most important variables (concomitant treatment) potentially affecting CTC detection. Although at meta-regression, no covariate significantly correlated with HR estimates, the analysis was not characterized by an adequate type-β error, which does not allow us to rule out that one or more covariates do represent a source of heterogeneity. As a corollary, these considerations should prompt investigators to validate the prognostic power of CTC by studying large homogeneous series of patients enrolled in multicenter prospective studies adopting standardized technical protocols.

The favorable results regarding survival analysis are at least in part conflicting with the observed CTC positivity rates per stage. The remarkable discrepancy between the CTC detection rates (32-41.7%) and the expected 5-year mortality rates (~10-33%) in patients with early stage melanoma (stage I, II) suggests that the false-positive rate might be an important hurdle to be tackled to definitely show CTC prognostic value. With regard to PCR-based methods, contamination events, amplification of pseudogenes, and illegitimate transcription can lead to false-positive results. Quantitative methods (both PCR-based and cytometric) can be used to set cutoff values distinguishing illegitimate transcription from marker expression by CTC (54). Furthermore, a broader implementation of cytometric methods, which allow for a morphologic identification of intact CTC, could significantly minimize the issue of false positivity associated with PCR-based methods.

Besides technical issues, the substantially unknown biology of CTC might contribute to the false-positive rate. In particular, the phenomenon of metastatic inefficiency can undermine the prognostic power of a strategy simply based on the presence/absence of CTC.

Some investigators (34, 56, 58) proposed the use of a cutoff to distinguish low-load from high-load CTC detection, which should distinguish low-risk from high-risk patients. To further increase the specificity of CTC detection assay meant as a prognostic tool, another approach is based on the use of target genes/proteins involved in cancer cell survival and biological aggressiveness, which is not the case for virtually all tumor markers thus far used.

On the opposite side, the CTC detection rate observed in stage IV patients (47.4%) is lower than expected (~80-90%), indicating that currently used detection systems are not sufficiently sensitive. Poor quality source material might lead to false-negative results. In effect, quality control has not been clearly mentioned in many publications: by contrast, only rigid per sample controls (e.g., sample quality assessment by quantitation of housekeeping gene expression) can provide reliable and comparable results. Moreover, the fact that most patients with metastatic disease are on treatment can have profound effects on CTC detection and prognostic value (27, 35, 48, 50, 57, 58). Finally, intermittent shedding of CTC into the bloodstream and genomic instability of malignant cells can also lead to false-negative results. A logical way to deal with these issues is to obtain multiple samples from each patient and to analyze the expression of multiple markers, respectively, to increase the likelihood of detecting CTC. However, in our analysis, only the use of multiple markers was associated with higher CTC detection rates.

In conclusion, although thousands of patients with melanoma have been enrolled over the last decade, available evidence is not sufficient to conclude that circulating melanoma cells are a biomarker reliable enough to be clinically implemented in the therapeutic decision-making process. However, the results of the present work justify researchers and clinicians to further investigate the prognostic power of CTC that can only be definitely validated (or confuted) by dissecting the molecular biology of CTC and by conducting large multicenter prospective studies based on homogeneous series of patients and designed according to standardized protocols.

Acknowledgments

We thank Salvatore Correra for his technical help with the literature search.

References

Imaging, Diagnosis, Prognosis


Koyanagi K, O'Day SJ, Gonzalez R, et al. Serial monitoring of circulating melanoma cells during...


The Prognostic Value of Circulating Tumor Cells in Patients with Melanoma: A Systematic Review and Meta-analysis

Simone Mocellin, Dave Hoon, Alessandro Ambrosi, et al.


<table>
<thead>
<tr>
<th>Updated version</th>
<th>Access the most recent version of this article at: <a href="http://clincancerres.aacrjournals.org/content/12/15/4605">http://clincancerres.aacrjournals.org/content/12/15/4605</a></th>
</tr>
</thead>
<tbody>
<tr>
<td>Supplementary Material</td>
<td>Access the most recent supplemental material at: <a href="http://clincancerres.aacrjournals.org/content/suppl/2006/08/09/12.15.4605.DC1">http://clincancerres.aacrjournals.org/content/suppl/2006/08/09/12.15.4605.DC1</a></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cited articles</th>
<th>This article cites 64 articles, 23 of which you can access for free at: <a href="http://clincancerres.aacrjournals.org/content/12/15/4605.full.html#ref-list-1">http://clincancerres.aacrjournals.org/content/12/15/4605.full.html#ref-list-1</a></th>
</tr>
</thead>
<tbody>
<tr>
<td>Citing articles</td>
<td>This article has been cited by 14 HighWire-hosted articles. Access the articles at: /content/12/15/4605.full.html#related-urls</td>
</tr>
</tbody>
</table>

**E-mail alerts**  
Sign up to receive free email-alerts related to this article or journal.

**Reprints and Subscriptions**  
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

**Permissions**  
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.