EUROPEAN HEMATOLOGY ASSOCIATION
haematologica
Journal of the European Hematology Association

Minimal residual disease prior to allogeneic hematopoietic cell transplantation in acute myeloid leukemia: a meta-analysis

by Sarah A. Buckley, Brent L. Wood, Megan Othus, Christopher S. Hourigan, Celalettin Ustun, Michael A. Linden, Todd E. DeFor, Michele Malagola, Chloe Anthias, Veronika Valkova, Christopher G. Kanakry, Bernd Gruhn, Francesco Buccisano, Beth Devine, and Roland B. Walter

Haematologica 2017 [Epub ahead of print]

Citation: Buckley SA, Wood BL, Othus M, Hourigan CS, Ustun C, Linden MA, DeFor TE, Malagola M, Anthias C, Valkova V, Kanakry CG, Gruhn B, Buccisano F, Devine B, and Walter RB. Minimal residual disease prior to allogeneic hematopoietic cell transplantation in acute myeloid leukemia: a meta-analysis. Haematologica. 2017; 102:xxx
doi:10.3324/haematol.2016.159343

Minimal residual disease prior to allogeneic hematopoietic cell transplantation in acute myeloid leukemia: a meta-analysis

Running Head: Pre-HCT MRD and Post-HCT Outcomes in AML

Sarah A. Buckley,[1] Brent L. Wood,[2] Megan Othus,[3] Christopher S. Hourigan,[4] Celalettin Ustun,[5] Michael A. Linden,[6] Todd E. DeFor,[7] Michele Malagola,[8] Chloe Anthias,[9,10] Veronika Valkova,[11] Christopher G. Kanakry,[12,13] Bernd Gruhn,[14] Francesco Buccisano,[15] Beth Devine,[16-18] and Roland B. Walter[19-21]

[1]Hematology/Oncology Fellowship Program, University of Washington, Seattle, WA, USA; [2]Division of Hematopathology, Department of Laboratory Medicine, University of Washington, Seattle, WA, USA; [3]Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA; [4]Hematology Branch, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD, USA; [5]Division of Hematology, Oncology and Transplantation, Department of Medicine, University of Minnesota, Minneapolis, MN, USA; [6]Division of Hematopathology, Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, MN, USA; [7]Biostatistics and Bioinformatics Core, Masonic Cancer Center, University of Minnesota, Minneapolis, MN, USA; [8]Unit of Blood Diseases and Stem Cell Transplantation, University of Brescia, A.O. Spedali Civili, Brescia, Italy; [9]Anthony Nolan Research Institute, London, UK; [10]Royal Marsden Hospital, London, UK; [11]Institute of Haematology and Blood Transfusion, Prague, Czech Republic; [12]Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University School of Medicine, Baltimore, MD, USA; [13]Experimental Transplantation and Immunology Branch, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA; [14]Department of Pediatrics, Jena University Hospital, Jena, Germany; [15]Department of Hematology, Fondazione Policlinico Tor Vergata, Rome, Italy; [16]Pharmaceutical Outcomes Research and Policy Program, University of Washington, Seattle, WA, USA; [17]Department of Health Services, University of Washington, Seattle, WA, USA; [18]Department of Biomedical Informatics, University of Washington, Seattle, WA, USA; [19]Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA; [20]Department of Medicine, Division of Hematology, University of Washington, Seattle, WA, USA; [21]Department of Epidemiology, University of Washington, Seattle, WA, USA

Running Head: Meta-analysis of pre-transplant MRD in AML

Corresponding Author: Sarah A. Buckley, MD, 1100 Fairview Ave North, Seattle, WA 98109, USA. Ph: +1-206-667-6172. Fax: +1-206-667-6519. Email: buckleys@uw.edu

Abstract: 210 words; Text: 3,056 words; Figure/Table count: 5/2; Supplemental Tables/Figures: 3/4; References: 39

Acknowledgements: S.A.B. is supported by a fellowship training grant from the National Heart, Lung, and Blood Institute/National Institutes of Health (NHLBI/NIH: T32-HL007093). R.B.W. is a Leukemia & Lymphoma Society Scholar in Clinical Research. This work was supported in part by the Intramural Research Programs of the National Heart, Lung, and Blood Institute and the National Cancer Institute of the National Institutes of Health.

ABSTRACT

Minimal residual disease prior to allogeneic hematopoietic cell transplantation has been associated with increased risk of relapse and death in patients with acute myeloid leukemia, but detection methodologies and results vary widely. We performed a systematic review and meta-analysis evaluating the prognostic role of minimal residual disease detected by polymerase chain reaction or multiparametric flow cytometry before transplant. We identified 19 articles published between January 2005 and June 2016 and extracted hazard ratios for leukemia-free survival, overall survival, and cumulative incidences of relapse and non-relapse mortality. Pre-transplant minimal residual disease was associated with worse leukemia-free survival (HR=2.76 [1.90-4.00]), overall survival (HR=2.36 [1.73-3.22]), and cumulative incidence of relapse (HR=3.65 [2.53-5.27]), but not non-relapse mortality (HR=1.12 [0.81-1.55]). These associations held regardless of detection method, conditioning intensity, and patient age. Adverse cytogenetics was not an independent risk factor for death or relapse. There was more heterogeneity among studies using flow cytometry-based than *WT1* polymerase chain reaction-based detection (I^2=75.1% vs. <0.1% for leukemia-free survival, 67.8% vs. <0.1% for overall survival, and 22.1% vs. <0.1% for cumulative incidence of relapse). These results demonstrate a strong relationship between pre-transplant minimal residual disease and post-transplant relapse and survival. Outcome heterogeneity among studies using flow-based methods may underscore site-specific methodological differences or differences in test performance and interpretation.

INTRODUCTION

Morphologic complete remission (CR), defined by the presence of <5% bone marrow blasts and recovery of peripheral blood counts, is the long-standing standard for response assessment in acute myeloid leukemia (AML).[1-5] Based on estimates of normal marrow cellularity,6 however, this cutoff allows for the presence of up to 10^{10} leukemic blasts or more. It is therefore not surprising that relapse remains the major cause of treatment failure among patients who have achieved a morphologic CR.[4, 5] Significant effort has gone into developing tools to identify minimal (or, perhaps more appropriately, measurable) residual disease (MRD), including multi-parametric flow cytometry (MFC) to enumerate myeloid cell populations with immunophenotypic abnormalities, polymerase chain reaction (PCR) to quantify leukemia-associated mutations or RNA transcript levels, and cytogenetic / fluorescence in situ hybridization to detect chromosome-level changes specific to the malignant clone. Among these modalities, MFC- and PCR-based approaches have the highest sensitivity and are increasingly employed in the clinic.[7-12]

A large number of studies has demonstrated worse outcomes for patients who have MRD compared to similarly treated patients in whom no MRD can be detected. This relationship has been observed during/after induction and post-remission chemotherapy courses as well as before and after hematopoietic cell transplantation (HCT).[7-12] The magnitude of the association between MRD status and risk of relapse varies widely between studies, however, as do the details of the detection methods. In addition to differences in the specifics of the MRD techniques across institutions, there are also differences in cut-points chosen to define MRD positivity, the patient material that is used to perform the MRD assay on (i.e., peripheral blood or bone marrow), and the timing as well as frequency with which MRD assessments are obtained. In this meta-analysis, we focused on MRD assessed immediately before allogeneic HCT in patients

with AML other than acute promyelocytic leukemia (APL). Besides ascertaining the relationship between pre-HCT MRD and post-transplant outcomes, we also investigated whether, and to what degree, the prognostic role of MRD is influenced by the method of MRD detection.

METHODS

We searched PubMed/MEDLINE and EMBASE (**Supplemental Table 1**) for English-language articles published between January 2005 and June 2016 that reported on the association between pre-HCT MRD (by PCR and/or MFC) and post-HCT survival in patients with non-APL AML in morphologic CR. Two authors (S.A.B. and R.B.W.) independently reviewed search results. We excluded studies with <15 patients or <6 months of follow-up. If needed, authors of included studies were contacted for additional information. Our search yielded 344 reports, which were screened according to 2009 PRISMA Guidelines (**Figure 1**). For studies of interest, we collected data on the number of patients, median/range age, median follow-up time, percent of patients with adverse risk cytogenetics (using the classification criteria reported by study authors), percent of patients receiving myeloablative (MA) vs. reduced intensity conditioning (RIC), interval between MRD detection and HCT, and details of the MRD detection method. We assessed risk of bias using an instrument based on the Quality in Prognostic Studies (QUIPS)[13] modified to reflect our judgment about potential biases (**Supplemental Table 2**). Finally, we obtained data for leukemia-free survival (LFS), overall survival (OS), and cumulative incidence of relapse (CIR) and non-relapse mortality (NRM) from the date of HCT. We used a hierarchical approach[14] to compare outcomes of MRD^{pos} and MRD^{neg} subjects: (i) When available, we used observed hazard ratios (HRs) and confidence intervals (CIs); (ii) When Kaplan-Meier curves were provided, we used Enguage Digitizer version 4.1 (www.markummitchell.github.io/engauge-digiitizer/) to calculate HRs and CIs based on an established algorithm;[15] (iii) For articles providing survival data at single time-points, we estimated HRs based on exponential decay.

We performed a random-effects meta-analysis, with inter-study heterogeneity described using the I^2 statistic[16] (STATA version 14; Stata Corp, College Station, TX). Cut-points

between MRD positivity and negativity were based on criteria specified by the individual publications. In one,[17] no cut-point was specified for Wilms Tumor 1 (*WT1*) transcript level. As other studies used cut points in the range of 50-70 copies/10^4 reference gene copies,[18-20] and as no events were observed at *WT1* levels <65, a cut-off of 70 was used. For another study[18] that used a *WT1* cutoff of 50, there were no relapses in the MRD^{neg} group (n=25) by 6.6 years. As no HR could be calculated, this study was not incorporated into pooled CIR results. For two studies in which HRs were extracted from survival curves,[21, 22] curves were portrayed for subgroups within MRD^{pos} and MRD^{neg} patients; here, a weighted average of the HR between groups by number of patients per group was used to obtain a final HR. For one study[19] reporting results by MFC and by *WT1* PCR, we used MFC results for overall analysis, as these data were more complete.

Subgroup analyses involved stratification by MRD detection method, age, and conditioning intensity. We calculated the ratio of the percentage of patients with adverse cytogenetics in the MRD^{pos} and MRD^{neg} groups. If HRs for survival outcomes were higher in studies where this ratio was greater, it would indicate that adverse cytogenetics might be an independent negative prognostic factor.[23] We used meta-regression to test this hypothesis.

RESULTS

Included studies

Our search yielded 19 unique publications with a total of 1,431 patients (**Table 1**).[17-19, 21, 22, 24-37] Details of transplant and conditioning regimens are shown in **Supplemental Table 3.** The sole method of MRD detection was MFC in 9 studies[22, 24, 26-29, 33, 36, 37] and *WT1* PCR in 5,[17, 18, 30-32] while one study reported results separately for MFC- and *WT1* PCR-based detection.[19] Four studies used combination methods;[21, 25, 34, 35] all of these included MFC, and 3 also included PCR-based detection. Among studies using MFC-based detection, the cut point between MRD positivity and negativity was fairly uniform: 11 of 14 used the limit of detection for the assay (around 0.1%), while 3 specified a cut-off of 0.1%,[26, 33, 36] which roughly corresponded to the limit of detection in these cases. In other words, heterogeneity in cut points was primarily determined by differences in performance characteristics and interpretation of the assay rather than cut points selected. Among studies that only used PCR-based methods, all assessed quantitative PCR for *WT1*, while one study[31] utilized a panel of other genes in addition to *WT1*. Two studies, both using combination approaches for MRD detection, targeted PCR at AML-specific mutations (e.g., Fms related tyrosine kinase 3 internal tandem duplication [*FLT3/ITD*])[21] or fusions genes (e.g., *RUNX1/RUNX1T1*)[25] present at diagnosis. Among studies quantifying *WT1* transcript levels, most normalized against expression of *ABL1*; MRD^{pos} cut-off levels varied between 50-70 copies *WT1* per 10^4 copies of *ABL1*.[17, 19, 32, 34]

Five studies were considered high risk of bias: MRD measurement technique was implicated in all cases, and study confounding was felt possible in 2 of these cases (**Figure 2**). For 11 studies, we were able to obtain HRs for all reported outcomes from the manuscript or personal communication; for the other 8 studies, HRs were extrapolated from Kaplan Meier curves or survival point estimates (n=4).[18, 26, 34, 36] MRD

was measured within 60 days of HCT in all studies in which this information was reported, and within 30 days in all but one study.[35]

Association between pre-HCT MRD status and post-HCT outcomes

Overall, MRD positivity was associated with worse LFS (HR=2.76 [1.90-4.00], I^2=70.0%), OS (HR=2.36 [1.73-3.22], I^2=59.7%), and CIR (HR=3.65 [2.53-5.27], I^2=37.9%) but not NRM (HR=1.12 [0.81-1.55], I^2<0.1%). After removing studies with high risk of bias in any domain, MRD remained strongly associated with worse LFS (HR=3.24 [2.17-4.83], I^2=64.5%), OS (HR=2.64 [1.87-3.72], I^2=57.8%), and CIR (HR=4.06 [2.70-6.12], I^2=48.0%) while, again, there was no statistically significant association with NRM (HR=1.18 [0.80-1.75] I^2=0.9%).

Effect of MRD detection method on post-HCT outcomes

In subgroup analyses, being MRD^{pos} was associated with increased risk of relapse and mortality regardless of the detection method (**Table 2**). For CIR, the HR for *WT1* PCR-based methods was statistically significantly larger than for MFC-based methods. **Figure 3** shows a forest plot for the 17 studies reporting on the primary outcome of LFS, while similar plots for OS, CIR, and NRM can be found in **Supplemental Figures 1-3**. Results for studies using MFC-based methods were more heterogeneous than those using *WT1* PCR or combination methods for LFS (I^2=75.1% vs. <0.1% and 57.2%), OS (I^2=67.8% vs. <0.1% and 12.5%), and CIR (I^2=22.1% vs. <0.1% and 6.7%). After excluding studies with a high risk of bias in any domain, *WT1* PCR-based studies and combination methods continued to have low heterogeneity for LFS, OS, and CIR (all I^2<0.1%), whereas MFC-based studies showed persistent and considerable heterogeneity for LFS (I^2=81.5%), OS (I^2=73.8%), and CIR (I^2=46.4%).

While all MFC-based studies analyzed bone marrow tissue, *WT1* PCR-based studies were mixed between use of marrow and peripheral blood for analysis. Restriction to studies that reported data from peripheral blood[18, 31, 32] yielded essentially identical results. Outcomes for MFC-based studies were similar regardless of whether residual disease was detected via gaiting for the original leukemia-associated immunophenotype or based on detecting a phenotype different from normal, although results for the latter were more heterogeneous (I^2 89.7% vs. 32.5% for LFS, 88.3% vs. 0.0% for OS, and 70.8% vs. 0.0% for CIR). There were no significant differences in outcomes between MFC-based studies by number of fluorochromes (<6 vs. ≥6) used.

Effect of patient age on post-HCT outcomes

On subgroup analysis of age 0-20,[18, 25, 26, 30, 36] 21-40,[28, 31, 34] and >40,[17, 19, 21, 22, 24, 27, 29, 32, 33, 35] we found no difference in the effect of MRD between groups. The same was true after exclusion of studies with high risk of bias. Among studies reporting on older patients, there was sufficient data to further stratify into ages 40-60 and >60 for the LFS endpoint; the HR for this outcome was similar in these subgroups (HR=2.67 [1.46-4.86], I^2=81.1%; HR=3.02 [0.90-10.08], I^2=52.3% respectively). When we restricted our analysis to studies using primarily MA conditioning, 2 were primarily pediatric (median age 0-20),[25, 30] 4 involved young adults (median age 20-40),[27, 28, 31, 34] and 5 involved older adults (median age >40).[19, 21, 24, 32, 35] The association between MRD and LFS was similar in all age groups among, though between-study heterogeneity was high (age 0-20: HR=3.45 [0.39-30.86], I^2=89.5%; age 20-40: HR=2.35 [1.10-5.02], I^2=72%; age >40: HR=3.56 [1.79-7.05], I^2=77.5%).

Effect of conditioning intensity on post-HCT outcomes

Next, we considered whether differences in conditioning intensity might explain between-study heterogeneity, particularly in light of conflicting results from Ustun et al[27] showing in a large cohort (n=203) that MA conditioning could compensate for the increased hazard for relapse and mortality associated with being MRD^{pos}, while Walter et al[29] showed no such effect in 241 patients. Among studies reporting LFS as an outcome, 14 reported on the fraction of MRD^{pos} and MRD^{neg} patients who underwent MA versus RIC HCT. To test whether higher intensity transplant might reduce the negative impact of being MRD^{pos}, we specifically analyzed studies in which >75% of MRD^{pos} and MRD^{neg} patients received MA HCT (n=12 for LFS endpoint) and compared results with studies where 0% of patients received MA HCT (n=3 for LFS endpoint). Results from Ustun at al[27] were reported separately for MA and RIC patients within their publication, and for the purposes of this analysis, we treated these sets of results as two separate studies. As shown in **Table 2** and as a forest plot in **Supplemental Figure 4**, there was no indication that MA conditioning was able to attenuate the negative effects associated with MRD positivity on LFS, OS, or CIR. In contrast, the HRs for MA studies were numerically higher than for the few RIC studies, although the large confidence intervals exclude a definitive conclusion as to whether conditioning intensity affects the association between MRD status and post-HCT outcomes. Exclusion of high-risk studies did not fundamentally change these results and conclusions. Not surprisingly, all studies using RIC conditioning involved older adults (the >40 age group as stratified above).

Effect of cytogenetic risk on post-HCT outcomes

Most studies reporting cytogenetics in MRD^{pos} and MRD^{neg} patients used Southwest Oncology Group[17, 27, 28, 32, 33] or 2010 Medical Research Council criteria,[21, 24, 29, 35] while one incorporated mutational profiling.[34] The ratio of the proportion of adverse risk cytogenetics among MRD^{pos} to MRD^{neg} patients ranged from roughly equal to 7.5 times

higher in the MRD^{pos} group. We used meta-regression to measure how HRs for LFS changed with variations in this risk ratio and found that differences in adverse risk cytogenetics between MRD^{pos} and MRD^{neg} groups did not account for a significant proportion of between-study variance (R^2): R^2 -9.15% (*P*=0.82) for all studies and -14.83% (*P*=0.92) after excluding high-risk studies (**Figure 4**). Results were similar when the study with the highest ratio of 7.5 was excluded from this analysis (*P*=0.62). Similarly, adverse risk cytogenetics was not an independent prognostic factor for OS (*P*=0.11), CIR (*P*=0.85), or NRM (*P*=0.99).

Testing for publication bias

Funnel plot analyses for each survival outcome are shown in **Figure 5** as a graph of log-HR versus the variance in the log-HR. These plots did not suggest a publication bias, although they indicated that the publication of studies considered to have high risk of bias could bias overall study results towards the null for LFS, OS, and CIR.

DISCUSSION

The findings from this meta-analysis support our major conclusion that the presence of MRD before allogeneic HCT identifies patients at higher risk of relapse and shorter survival relative to patients in whom no evidence of MRD is found. Although we were unable to incorporate results from one study with an incalculable HR for CIR (based on the lack of relapses among MRD^{neg} patients), the findings from that report similarly support our conclusion. The association between MRD and post-HCT relapse and mortality is robust and is seen within all patient ages and regardless of which detection method is used. It is similarly found in those undergoing MA conditioning as well as RIC transplants without discernible difference in strength of association between these cohorts, suggesting that higher conditioning intensity may not be able to overcome the adverse impact of MRD. To the extent that we were able to control for differences in cytogenetic risk with meta-regression, the negative impact of being MRD^{pos} superseded any potential adverse effects of having poor-risk cytogenetics. In comparison, our analysis indicates that pre-HCT MRD is not associated with a significantly increased risk of NRM, in line with the notion that the association between pre-HCT MRD and OS is entirely accounted for by disease relapse without significant contribution from HCT toxicity.

Although our meta-analysis demonstrates a significant association between pre-HCT MRD status and post-HCT outcomes with both *WT1* PCR- and MFC-based assays, we found a greater degree of heterogeneity in survival estimates in studies with MFC-based detection methods. This heterogeneity could not be accounted for by differences in patient age, conditioning intensity, or cytogenetic risk. In addition, the cut-points between MRD positivity and negativity were primarily determined by the limits of detection of each particular assay, indicating that chosen cut-points are unlikely to account for

heterogeneity. We were, however, able to show that at least some of this heterogeneity may be accounted for by study-specific differences in approach to MFC, with studies detecting residual disease based on initial leukemia-associated immunophenotypes having more uniform results than those using a “different from normal” approach. Other possible causes of heterogeneity might include site-specific differences in MFC methodology, including differences in antigens and fluorochromes used, methods for cell lysis, number of events collected, or specifics of the aspirate used for analysis, with increasing risk of hemodilution with each pull. If such speculation is correct, efforts towards standardization/harmonization of MRD methods – as pioneered in acute lymphoblastic leukemia[38] and currently underway for AML – might ultimately lead to less heterogeneous data with MFC-based MRD assays. In contrast, despite some heterogeneity in PCR targets and cut-points, PCR methodology may be relatively more standardized, accounting for more uniform results. As an illustration, the risk estimates for being MRD^{pos} by PCR-quantified *WT1* transcript levels are very similar across several studies, indicating that this method yields highly reproducible results for pre-HCT risk stratification. Even in the smallest studies,[17,19] in which there was no statistically significant relationship between MRD and LFS, observed HRs were consistent with the other, larger studies. One might wonder whether using more than one method to detect MRD might lead to more sensitive detection and stronger associations with relapse and survival, indicated by higher HRs. We found that studies using combination methods of MRD detection did not show stronger associations with survival outcomes over studies using either MFC- or *WT1* PCR-based methods. That said, all four of these ‘combinations’ involved MFC-based detection, and the heterogeneity within the combination group may simply underscore the heterogeneity in MFC-based studies as a whole. Alternatively, MFC and *WT1* PCR are both potentially highly sensitive tests, and using multiple modalities may not add much additional sensitivity in detection, or

increases in assay sensitivity beyond current limits may not lead to appreciably stronger associations with survival outcomes.

Although our studies highlight the importance of pre-HCT MRD, we were unable to account for inter-study differences in the selection of patients for HCT, which may impact post-HCT results. It is conceivable that different strategies in allocating patients to different post-remission treatment strategies could affect our study results. Given the nature of our analysis, we were only able to test the effects of select covariates and only in aggregate fashion. Similarly, we were not able to control for the considerable heterogeneity in transplant conditioning regimens, donor sources, graft characteristics, and immunosuppression, all of which could potentially influence relapse and death. Absent individual patient data, we were not able to assess whether higher levels of MRD were associated with higher risk of relapse. Regardless of these limitations, our results demonstrate a strong relationship between pre-HCT MRD status and post-HCT relapse and survival but not NRM. Further studies are needed to determine how pre-HCT MRD status should guide therapeutic decisions, either through treatment intensification for MRD^{pos} patients or possibly de-intensification for patients who are found to be MRD^{neg} by a reliable method.

ACKNOWLEDGEMENTS

Conflict of interest: C.S.H. receives research funding from SELLAS Life Sciences Group AG. M.A.L. is a consultant to Bristol Myers Squibb and receives royalties from Cell Signaling Technology.

Authorship: S.A.B. and R.B.W. were responsible for the concept of this meta-analysis and contributed to the literature search, data collection quality assessment, data analysis

and interpretation, and writing of the manuscript. B.L.W. provided guidance on the risk of bias assessment, analyzed and interpreted data, and revised the manuscript. M.O. and B.D. provided guidance on statistical methodology, analyzed and interpreted data, and revised the manuscript. C.S.H, C.U., M.A.L., T.E.D., M.M., C.A., V.V., C.G.K., B.G., and F.B. contributed data, analyzed and interpreted data, and revised the manuscript.

References:

1. Bisel HF. Criteria for the evaluation of response to treatment in acute leukemia. Blood. 1956;11:676-677.
2. Cheson BD, Bennett JM, Kopecky KJ, et al. Revised recommendations of the International Working Group for Diagnosis, Standardization of Response Criteria, Treatment Outcomes, and Reporting Standards for Therapeutic Trials in Acute Myeloid Leukemia. J Clin Oncol. 2003;21(24):4642-4649.
3. Döhner H, Estey EH, Amadori S, et al. Diagnosis and management of acute myeloid leukemia in adults: recommendations from an international expert panel, on behalf of the European LeukemiaNet. Blood. 2010;115(3):453-474.
4. Ferrara F, Schiffer CA. Acute myeloid leukaemia in adults. Lancet. 2013;381(9865):484-495.
5. Döhner H, Weisdorf DJ, Bloomfield CD. Acute Myeloid Leukemia. N Engl J Med. 2015;373(12):1136-1152.
6. Harrison WJ. The total cellularity of the bone marrow in man. J Clin Pathol. 1962;15:254-259.
7. Coustan-Smith E, Campana D. Should evaluation for minimal residual disease be routine in acute myeloid leukemia? Curr Opin Hematol. 2013;20(2):86-92.
8. Hourigan CS, Karp JE. Minimal residual disease in acute myeloid leukaemia. Nat Rev Clin Oncol. 2013;10(8):460-471.
9. Grimwade D, Freeman SD. Defining minimal residual disease in acute myeloid leukemia: which platforms are ready for "prime time"? Blood. 2014;124(23):3345-3355.
10. Paietta E. Minimal residual disease in acute myeloid leukemia: coming of age. Hematology Am Soc Hematol Educ Program. 2012;2012:35-42.
11. Hokland P, Ommen HB, Mulé MP, Hourigan CS. Advancing the minimal residual disease concept in acute myeloid leukemia. Semin Hematol. 2015;52(3):184-192.
12. Ommen HB. Monitoring minimal residual disease in acute myeloid leukaemia: a review of the current evolving strategies. Ther Adv Hematol. 2016;7(1):3-16.
13. Hayden JA, van der Windt DA, Cartwright JL, Cote P, Bombardier C. Assessing bias in studies of prognostic factors. Ann Intern Med. 2013;158(4):280-286.
14. Broglio KR, Quintana M, Foster M, et al. Association of pathologic complete response to neoadjuvant therapy in HER2-positive breast cancer with long-term outcomes: a meta-analysis. JAMA Oncol. 2016;2(6):751-760.
15. Tierney JF, Stewart LA, Ghersi D, Burdett S, Sydes MR. Practical methods for incorporating summary time-to-event data into meta-analysis. Trials. 2007;8:16.
16. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med. 2002;21(11):1539-1558.
17. Candoni A, Toffoletti E, Gallina R, et al. Monitoring of minimal residual disease by quantitative WT1 gene expression following reduced intensity conditioning allogeneic stem cell transplantation in acute myeloid leukemia. Clin Transplant. 2011;25(2):308-316.
18. Jacobsohn DA, Tse WT, Chaleff S, et al. High WT1 gene expression before haematopoietic stem cell transplant in children with acute myeloid leukaemia predicts poor event-free survival. Br J Haematol. 2009;146(6):669-674.
19. Rossi G, Carella AM, Minervini MM, et al. Optimal time-points for minimal residual disease monitoring change on the basis of the method used in patients with acute myeloid leukemia who underwent allogeneic stem cell transplantation: a comparison between multiparameter flow cytometry and Wilms' tumor 1 expression. Leuk Res. 2015;39(2):138-143.

20. Olszewski M, Chou PM, Huang W, Tallman S, Kletzel M. Correlation of minimal residual disease by assessing Wilms tumor gene expression and engraftment by variable number of tandem repeats in children with leukemia posthematopoietic stem cell transplantation. Pediatr Dev Pathol. 2006;9(3):203-209.
21. Kanakry CG, Tsai HL, Bolanos-Meade J, et al. Single-agent GVHD prophylaxis with posttransplantation cyclophosphamide after myeloablative, HLA-matched BMT for AML, ALL, and MDS. Blood. 2014;124(25):3817-3827.
22. Anthias C, Dignan FL, Morilla R, et al. Pre-transplant MRD predicts outcome following reduced-intensity and myeloablative allogeneic hemopoietic SCT in AML. Bone Marrow Transplant. 2014;49(5):679-683.
23. Buccisano F, Maurillo L, Spagnoli A, et al. Cytogenetic and molecular diagnostic characterization combined to postconsolidation minimal residual disease assessment by flow cytometry improves risk stratification in adult acute myeloid leukemia. Blood. 2010;116(13):2295-2303.
24. Araki D, Wood BL, Othus M, et al. Allogeneic hematopoietic cell transplantation for acute myeloid leukemia: time to move toward a minimal residual disease-based definition of complete remission? J Clin Oncol. 2016;34(4):329-336.
25. Zheng C, Zhu X, Tang B, et al. The impact of pre-transplant minimal residual disease on outcome of intensified myeloablative cord blood transplant for acute myeloid leukemia in first or second complete remission. Leuk Lymphoma. 2016;57(6):1398-1405.
26. Bleyzac N, Cuzzubbo D, Renard C, et al. Improved outcome of children transplanted for high-risk leukemia by using a new strategy of cyclosporine-based GVHD prophylaxis. Bone Marrow Transplant. 2016;51(5):698-704.
27. Ustun C, Courville EL, DeFor T, et al. Myeloablative, but not reduced-intensity, conditioning overcomes the negative dffect of flow-cytometric evidence of leukemia in acute myeloid leukemia. Biol Blood Marrow Transplant. 2016;22(4):669-675.
28. Tian H, Chen GH, Xu Y, et al. Impact of pre-transplant disease burden on the outcome of allogeneic hematopoietic stem cell transplant in refractory and relapsed acute myeloid leukemia: a single-center study. Leuk Lymphoma. 2015;56(5):1353-1361.
29. Walter RB, Gyurkocza B, Storer BE, et al. Comparison of minimal residual disease as outcome predictor for AML patients in first complete remission undergoing myeloablative or nonmyeloablative allogeneic hematopoietic cell transplantation. Leukemia. 2015;29(1):137-144.
30. Woehlecke C, Wittig S, Arndt C, Gruhn B. Prognostic impact of WT1 expression prior to hematopoietic stem cell transplantation in children with malignant hematological diseases. J Cancer Res Clin Oncol. 2015;141(3):523-529.
31. Goswami M, McGowan KS, Lu K, et al. A multigene array for measurable residual disease detection in AML patients undergoing SCT. Bone Marrow Transplant. 2015;50(5):642-651.
32. Valkova V, Polak J, Markova M, et al. Minimal residual disease detectable by quantitative assessment of WT1 gene before allogeneic stem cell transplantation in patients in first remission of acute myeloid leukemia has an impact on their future prognosis. Clin Transplant. 2013;27(1):E21-29.
33. Bastos-Oreiro M, Perez-Corral A, Martinez-Laperche C, et al. Prognostic impact of minimal residual disease analysis by flow cytometry in patients with acute myeloid leukemia before and after allogeneic hemopoietic stem cell transplantation. Eur J Haematol. 2014;93(3):239-246.
34. Wang Y, Liu DH, Liu KY, et al. Impact of pretransplantation risk factors on post transplantation outcome of patients with acute myeloid leukemia in remission after haploidentical hematopoietic stem cell transplantation. Biol Blood Marrow Transplantat. 2013;19(2):283-290.

35. Grubovikj RM, Alavi A, Koppel A, Territo M, Schiller GJ. Minimal residual disease as a predictive factor for relapse after allogeneic hematopoietic stem cell transplant in adult patients with acute myeloid leukemia in first and second complete remission. Cancers (Basel). 2012;4(2):601-617.
36. Leung W, Pui CH, Coustan-Smith E, et al. Detectable minimal residual disease before hematopoietic cell transplantation is prognostic but does not preclude cure for children with very-high-risk leukemia. Blood. 2012;120(2):468-472.
37. Laane E, Derolf AR, Bjorklund E, et al. The effect of allogeneic stem cell transplantation on outcome in younger acute myeloid leukemia patients with minimal residual disease detected by flow cytometry at the end of post-remission chemotherapy. Haematologica. 2006;91(6):833-836.
38. van Dongen JJ, van der Velden VH, Bruggemann M, Orfao A. Minimal residual disease diagnostics in acute lymphoblastic leukemia: need for sensitive, fast, and standardized technologies. Blood. 2015;125(26):3996-4009.
39. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ. 2003;327(7414):557-560.

Table 1. Characteristics of included studies

Study	MRD method	MRD source	Cutoff for MRD^{pos}	MRD^{neg} (n)	MRD^{pos} (n)	Age, median (range)	% MA
Bleyzac et al[26]	MFC (LAIP)	BM	0.1%	18	14	9 (0-19)	NR
Ustun et al[27]	MFC (DFN, 4-color)	BM	Limit of detection (0.1%)	178	25	47 (0-74)	39% MRD^{neg} 60% MRD^{pos}
Zheng et al[25]	MFC (LAIP, 4-color) or PCR (fusion genes, multiple)	BM	MFC: 0.01% PCR: limit of detection	40	32	MRD^{neg} 16 (3-28) MRD^{pos} 19 (6-36)	100%
Araki et al[24]	MFC (DFN, 10-color)	BM	Limit of detection (0.1%)	235	76	MRD^{neg} 47 (19-71) MRD^{pos} 51 (18-72)	100%
Goswami et al[31]	PCR (*WT1*, multi-gene)	PB	Different for each gene	38	10	MRD^{neg} 34 (12-59) MRD^{pos} 34 (16-53)	89% MRD^{neg} 100% MRD^{pos}
Rossi et al[19]	MFC (LAIP, 6-color) PCR (*WT1*)	BM	MFC: 0.1% *WT1*: 64 / 10^4 copies ABL	22 (MFC) 19 (PCR)	8 (MFC) 10 (PCR)	44 (18-64)	100%
Tian et al[28]	MFC (LAIP, 4-color)	BM	Limit of detection	21	32	MRD^{neg} 31 (15-55) MRD^{pos} 32 (16-58)	NR

Walter et al[29]	MFC (DFN, 10-color)	BM	Limit of detection (0.1%)	65	21	MRD^{neg} 62 (20-75) MRD^{pos} 63 (33-74)	0%
Woehlecke et al[30]	PCR (*WT1*)	PB or BM	5×10^{-3} normalized to *β2M* expression	17	23	MRD^{neg} 4 (1-21) MRD^{pos} 13 (2-18)	100%
Anthias et al[22]	MFC (LAIP, 3-color)	BM	Limit of detection (0.4%)	53	35	MRD^{neg} 44 (18-70) MRD^{pos} 52 (21-70)	40% MRD^{neg} 60% MRD^{pos}
Bastos-Oriero et al[33]	MFC (LAIP, 4-color)	BM	0.1%	18	11	MRD^{neg} 41 (19-62) MRD^{pos} 50 (19-63)	100% MRD^{neg} 72% MRD^{pos}
Kanakry et al[21]	MFC (LAIP), PCR (*FLT3*, *NPM1*), and/or cytogenetics / FISH	BM	MFC: limit of detection PCR: limit of detection	76	25	51 (20-66)	100%
Wang et al[34]	MFC (LAIP) and PCR (*WT1*)	BM	MFC: limit of detection *WT1*: 60 / 10^4 copies ABL	110	20	26 (3-54)	100%
Grubovikj et al[35]	MFC (DFN) or cytogenetics / FISH		Limit of detection	40	19	MRD^{neg} 43 (20-65) MRD^{pos} 50 (28-65)	90% MRD^{neg} 84% MRD^{pos}

Leung et al[36]	MFC (LAIP, 4-color)	BM	0.1%	27	9	(Pediatric)	100%
Valkova et al[32]	PCR (*WT1*)	PB	50 / 10^4 copies *ABL1*	29	13	MRD^{neg} 43 (20-63) MRD^{pos} 51 (36-63)	79% MRD^{neg} 85% MRD^{pos}
Candoni et al[17]	PCR (*WT1*)	BM	70 / 10^4 copies *ABL1*	5	13	MRD^{neg} 61 (39-66) MRD^{pos} 61 (36-68)	0%
Jacobsohn et al[18]	PCR (*WT1*)	PB	0.5 (normalized to *WT1* level in control cells)	25	11	10 (3-22)	100%
Laane et al[37]	MFC (LAIP, 3-color)	BM	Limit of detection	12	5	(Adult)	100%

Abbreviations: PB, peripheral blood; BM, bone marrow; NR, not reported; MFC, multiparametric flow cytometry; LAIP, leukemia-associated immunophenotype; MA, myeloablative; DFN, different from normal; *WT1*, Wilms tumor 1; *β2M*, β2-microglobulin; FISH, fluorescence in situ hybridization; *FLT3*, Fms related tyrosine kinase 3

Table 2: Pooled HRs [95% CI] and inter-study heterogeneity for all studies (above) and excluding high risk of bias (below). Only fields pooled from ≥2 studies are reported; otherwise, fields are left blank. Colored boxes indicate degree of heterogeneity as defined by the I^2 statistic: 0-24.9% = low (□), 25-75% = moderate (■), 75.1-100% = high (■).39 Cells are filled only if two or more studies contribute to the analysis.

All Studies				
Subset	**OS**	**LFS**	**CIR**	**NRM**
Method				
MFC	1.98 [1.26-3.10] ■	2.41 [1.36-4.29] ■	2.81 [1.94-4.08] □	1.11 [0.63-1.95] □
PCR	5.25 [3.08-8.95] □	5.80 [3.57-9.42] □	9.53 [4.48-20.29] □	1.51 [0.57-4.00] □
Combination	1.86 [1.25-2.77] □	1.79 [1.06-3.01] ■	3.73 [1.94-7.18] □	1.15 [0.57-2.33] □
Median age				
0-20	3.12 [1.29-7.57] ■	3.33 [0.95-11.6] ■	3.57 [0.67-18.91] ■	1.13 [0.52-2.4] □
21-40	2.60 [1.36-4.99] ■	3.02 [1.27-7.16] ■	5.13 [2.37-9.64] □	--
>40	2.25 [1.47-3.47] ■	2.69 [1.64-4.42] ■	3.33 [2.18-5.11] ■	1.23 [0.77-1.97] □
Conditioning				
>75% MA	2.64 [1.77-3.93] ■	2.86 [1.80-4.55] ■	4.21 [2.70-6.58] ■	1.39 [0.94-2.07] □
0% MA	2.05 [0.78-5.39] ■	2.09 [1.33-3.29] □	3.23 [1.88-5.53] □	0.58 [0.22-1.52] □

Excluding Studies with High Risk of Bias				
Subset	**OS**	**LFS**	**CIR**	**NRM**

Method				
MFC	2.19 [1.29-3.72] ■	2.77 [1.39-5.50] ■	2.90 [1.81-4.64] ■	1.11 [0.63-1.95] □
PCR	4.60 [2.60-8.14] □	5.14 [3.04-8.72] □	9.53 [4.48-20.29] □	1.28 [0.41-4.03] ■
Combination	2.57 [1.52-4.33] □	2.81 [1.70-4.66] □	4.53 [2.30-8.92] □	
Median age				
0-20	4.41 [1.65-11.8] ■	5.89 [1.90-18.2] ■	--	1.16 [0.18-7.58] ■
21-40	3.29 [1.39-7.79] ■	4.13 [1.19-14.3] ■	5.66 [2.80-11.4] □	--
>40	2.46 [1.56-3.86] ■	3.06 [1.85-5.05] ■	3.33 [2.18-5.11] ■	1.23 [0.77-1.97] □
Conditioning				
>75% MA	3.39 [2.20-5.22] ■	4.09 [2.53-6.62] ■	4.72 [2.97-7.50] ■	1.42 [0.90-2.25] □
0% MA	2.05 [0.78-5.39] ■	2.09 [1.33-3.29] □	3.23 [1.88-5.53] □	0.58 [0.22-1.52] □

Figure Legends:

Figure 1: PRISMA flow diagram for study selection

Figure 2: Risk of bias assessment illustrating review authors' judgments about each risk of bias item for each included study.

Figure 3: Forest plot showing hazard ratio (effect size, ES) for leukemia-free survival with pooling of results for each minimal residual disease detection method. Columns indicate study size (N) and whether each study carries a high risk of bias (Bias Risk). Within groups, studies are listed by year of publication.

Figure 4: Meta-regression analysis showing the effect of the ratio of percent of MRD^{pos} patients with adverse cytogenetics to percent of MRD^{neg} patients with adverse cytogenetics on hazard for leukemia-free survival. A flat line indicates no relationship, and this is shown for all studies (a) and after excluding studies with high risk of bias (b).

Figure 5: Funnel plot analysis for survival outcomes. Shown are (a) leukemia-free survival, (b) overall survival, (c) cumulative incidence of relapse, (d) non-relapse mortality.

301 MEDLINE

43 EMBASE

344 considered

258 excluded based on title and abstract
- Non-human (n=9)
- Wrong study type (n=131)
- No MRD measurement (n=39)
- Wrong disease (n=30)
- Wrong timing of MRD measurement (n=41)
- No allogeneic HCT (n=8)

86 full papers considered

19 papers included

67 excluded based on full paper
- Wrong study type (n=2)
- No MRD measurement (n=11)
- Wrong disease (n=1)
- Wrong timing of MRD measurement (n=17)
- No allogeneic HCT (n=10)
- Not enough information (n=12)
- Too few patients (n=3)
- Redundant data (n=11)

Study	1	2	3
Bleyzac N (2016)	●	●	◐
Ustun C (2016)	○	○	○
Zheng C (2016)	●	○	○
Araki D (2015)	○	○	○
Goswami M (2015)	◐	◐	○
Rossi G (2015)	◐	◐	○
Tian H (2015)	●	○	◐
Walter RB (2015)	○	○	○
Woehlecke C (2015)	○	◐	○
Anthias C (2014)	○	○	○
Bastos-Oriero M (2014)	○	○	◐
Kanakry CG (2014)	●	○	○
Wang Y (2013)	◐	◐	◐
Grubovikj RM (2012)	○	○	○
Leung W (2012)	◐	◐	◐
Valkova V (2012)	◐	○	○
Candoni A (2011)	◐	○	○
Jacobsohn DA (2009)	◐	◐	◐
Laane E (2006)	●	●	◐

Risks of Bias

1 Prognostic factor measurement

2 Study confounding

3 Statistical analysis and reporting

Color Key

○ Low risk

◐ Moderate risk

● High risk

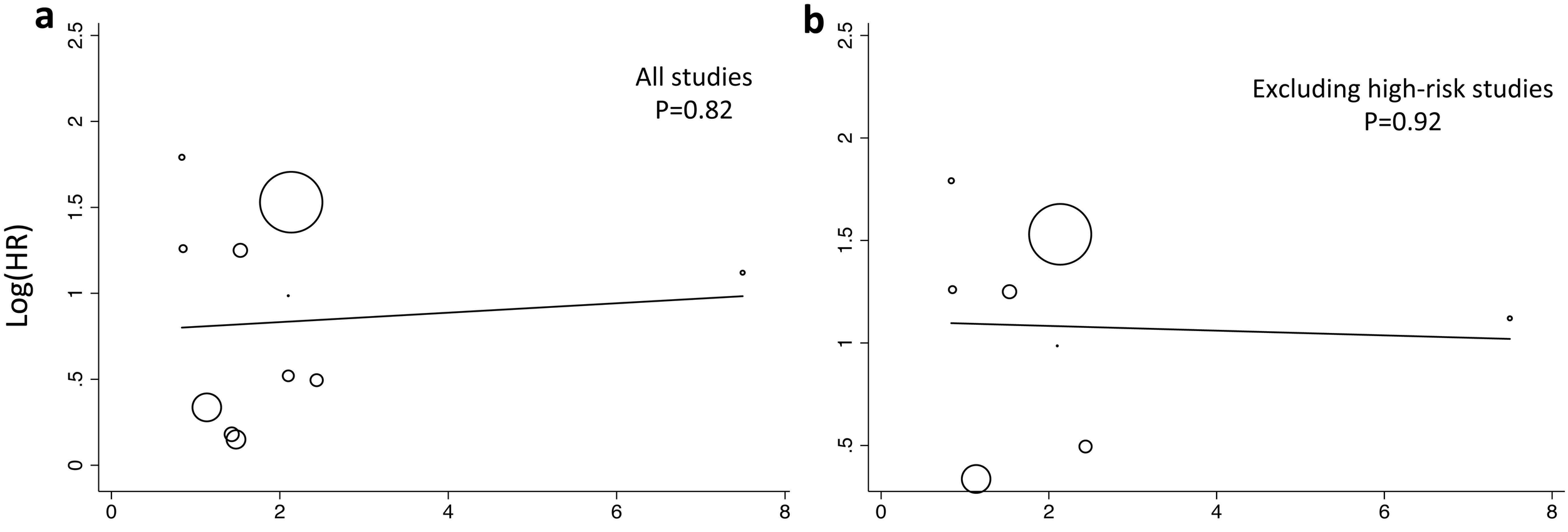
Impact of MRD on Leukemia-Free Survival

	Study	N	Bias Risk	ES [95% CI]
MFC	Ustun	203		1.40 (0.90, 2.40)
	Araki	331		4.62 (3.31, 6.44)
	Rossi	30		10.18 (1.95, 52.96)
	Walter	86		1.64 (0.90, 3.00)
	Tian	53	High	1.68 (0.77, 3.71)
	Bastos-Oriero	29		3.06 (0.87, 10.92)
	Laane	17	High	0.92 (0.08, 11.03)
	Subtotal (I-squared = 75.1%, p = 0.000)			2.41 (1.36, 4.29)
	.			
PCR	Goswami	48		8.08 (3.17, 20.55)
	Rossi	29		3.85 (0.85, 17.49)
	Woehlecke	40		11.23 (3.25, 38.82)
	Valkova	42		6.00 (1.93, 18.60)
	Candoni	18		2.68 (0.23, 30.66)
	Jacobsohn	36		3.53 (1.33, 9.25)
	Subtotal (I-squared = 0.0%, p = 0.660)			5.80 (3.57, 9.42)
	.			
Combination	Zheng	72	High	1.20 (0.59, 2.39)
	Kanakry	101	High	1.16 (0.63, 2.12)
	Wang	130		2.27 (1.12, 4.65)
	Grubovikj	59		3.49 (1.71, 7.11)
	Subtotal (I-squared = 57.2%, p = 0.072)			1.79 (1.06, 3.01)

NOTE: Weights are from random effects analysis

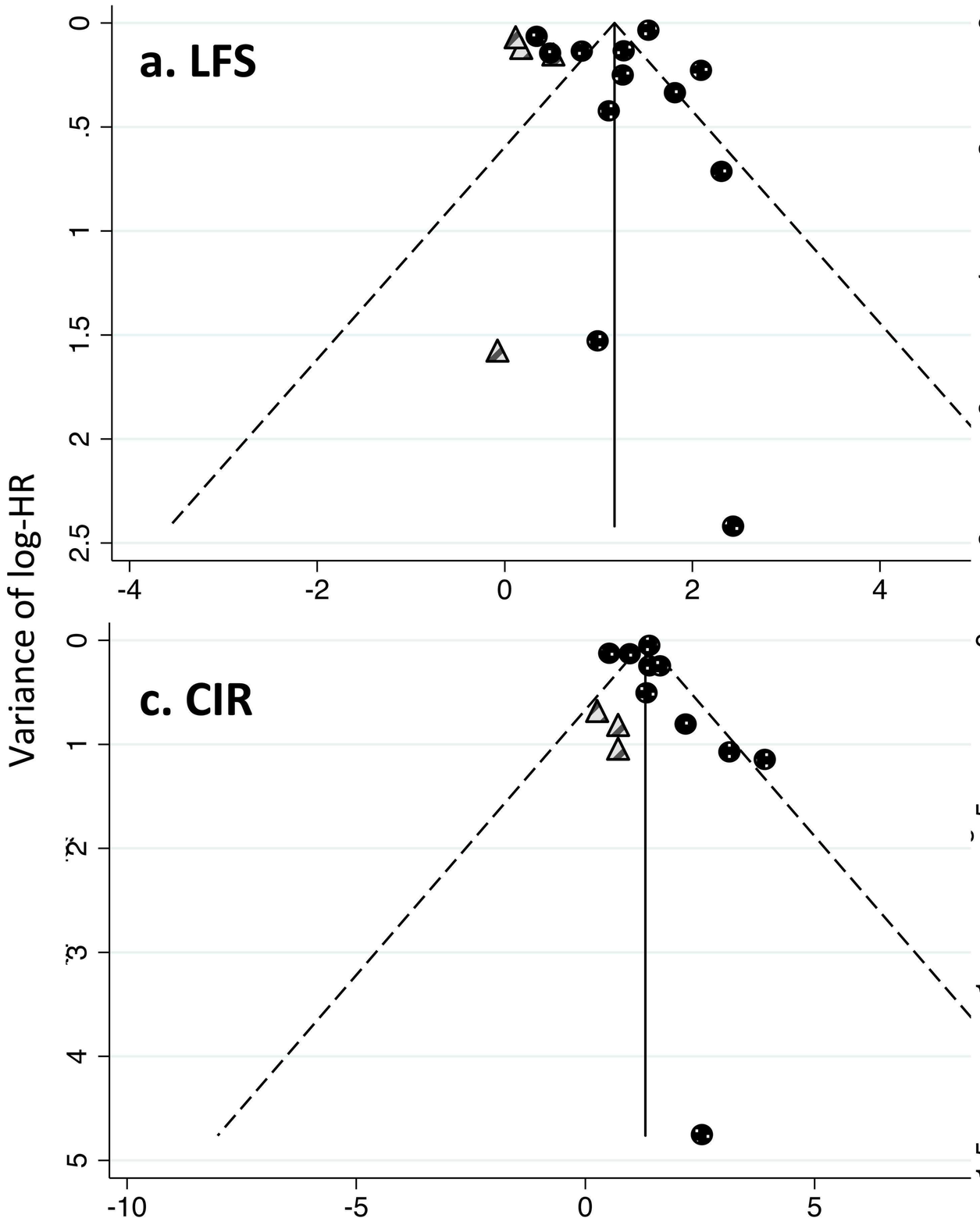
.05 .1 1 10 20

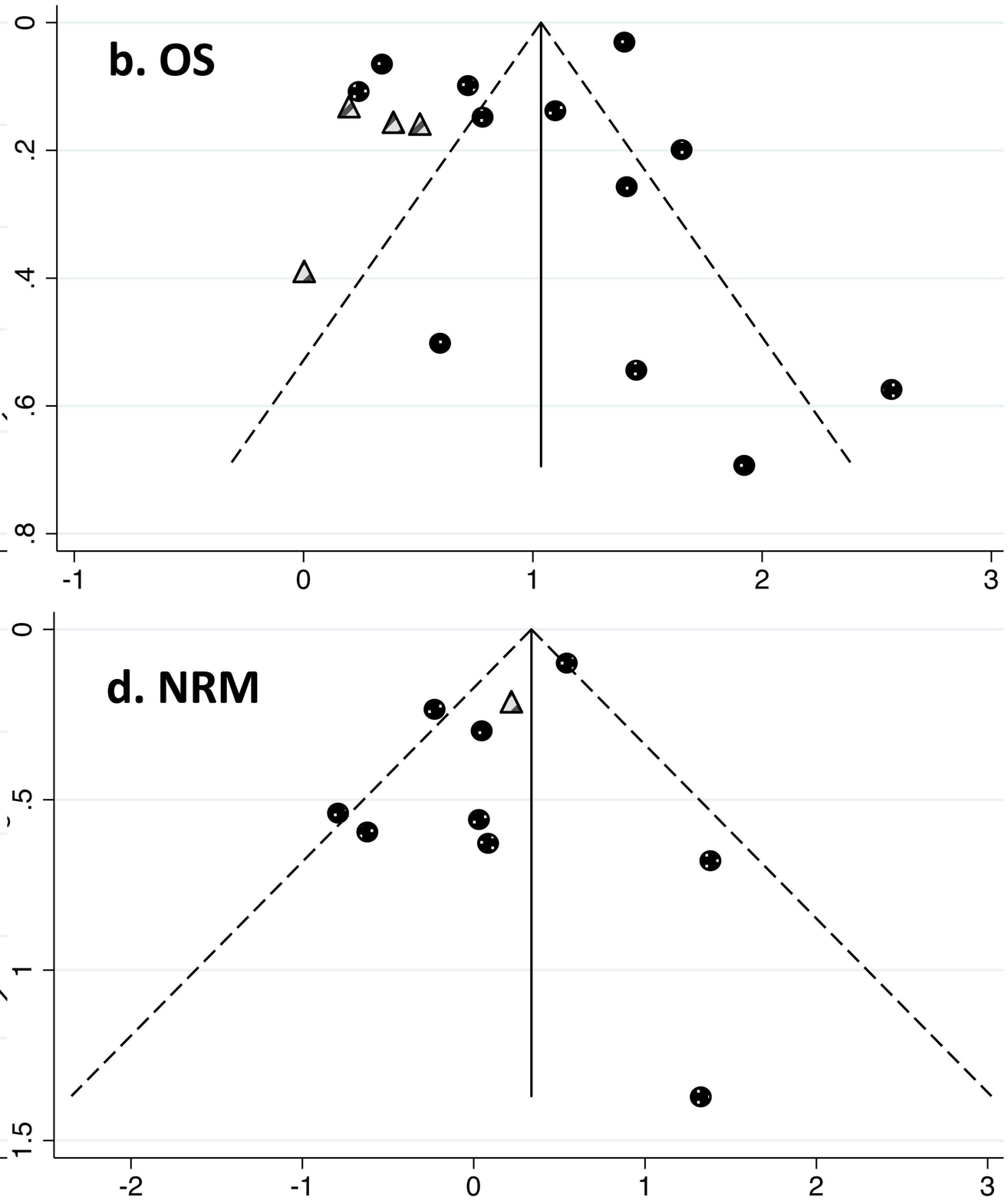
MRD+ status better MRD+ status worse



Ratio: MRD+ (percent adverse cytogenetics) / MRD- (percent adverse cytogenetics)

Funnel plots with pseudo 95% confidence limits





Log-HR

Low or moderate risk of bias High risk of bias

SUPPLEMENTAL MATERIALS

Supplemental Table 1: Search strategies for PubMed and EMBASE, with additional restrictions on year of publication starting January 1st, 2005 and English-language publications. In EMBASE, only published articles were included.

PubMed strategy	**EMBASE strategy**
(“acute myeloid leukemia” OR “acute myelogenous leukemia” OR “acute myeloid leukaemia” OR “acute myelogenous leukaemia” OR "acute myeloblastic leukemia" OR "actue myeloblastic leukaemia" OR AML OR "leukemia, myeloid, acute"[MeSH]) **AND** (transplant OR transplantation OR transplanted OR transplants OR transplantations OR HCT OR HSCT OR allogeneic OR BMT OR “stem cell transplantation”[MeSH]) **AND** (“residual disease” OR MRD OR "residual leukemia" OR "stringent complete remission" OR "stringent CR" OR “neoplasm, residual”[MeSH])	('acute myeloblastic leukemia'/exp OR "acute myeloid leukemia" OR "acute myeloid leukaemia" OR "acute myelogenous leukemia" OR "acute myelogenous leukaemia" OR "acute myeloblastic leukemia" OR "acute myeloblastic leukaemia" OR AML) **AND** ('minimal residual disease'/exp OR "residual cancer" OR "residual disease" OR "residual leukemia" OR "residual leukaemia" OR mrd OR “stringent complete remission” OR “stringent CR”) **AND** ('stem cell transplantation'/exp OR HSCT OR HCT OR transplant OR transplants OR transplanted OR transplantation OR allogeneic OR BMT)

Supplemental Table 2: Risk of bias assessment tool used to assign risk of bias

Bias Domains	Study Characteristics	Risk of Bias
Prognostic Factor Measurement	**Pre-HCT measurement of MRD is appropriate** (a) MRD detection method must be clearly described, valid, and reliable. (b) Continuous variables are reported or pre-specified cut points are used. (c) MRD is measured close enough to the start of transplant to capture a true "pre-transplant" state	**High risk**: MRD detection methods not described or likely to be inaccurate based on the following criteria: • MRD measured >60 days before transplant • For MFC, methodology is suspect based on (1) reported sensitivity not consistent with number of cells collected and reagent panels used, (2) details such as cells/tube and antibody panels neither provided nor referenced, (3) center has not had prior publications with referenced protocols if not using their own protocol, or (4) $<10^5$ cells/tube used. • For PCR, methodology is suspect based on (1) >24 hours between specimen collection and RNA extraction, (2) lack of negative and standard controls, and (3) failure to perform the assay with >1 replicate. **Moderate risk**: cut points between MRD+ and MRD- are chosen based on exploratory analysis without a validation cohort and/or time between measurement and transplant not reported. Further, there is insufficient information to assess bias in MFC or PCR methodology. **Low risk**: MRD measurement is valid, with a pre-specified cut-point between MRD+ and MRD-, and MRD is measured within 60 days of transplant.
Study Confounding	**Important potential confounding factors are described** (a) Confounders are measured across all participants and are reported separately for MRD+ and MRD- patients; key covariates are age, cytogenetics, and conditioning intensity. (b) Inclusion of patients not in CR may bias the MRD+ group toward worse outcomes.	**High risk**: no key covariates are reported for MRD+ and MRD- patients. **Moderate risk**: only some key covariates are reported for MRD+ and MRD- patients. **Low risk**: all key covariates are reported for MRD+ and MRD- patients.
Statistical Analysis and Reporting	**The statistical analysis is appropriate, and all primary outcomes are reported** (a) Statistical methods are described, and there is no selective reporting of results (b) Hazard ratios for outcomes should be accurate	**High risk**: the reported results are likely to be biased related to selective reporting of data (not all outcomes described in methods reported in results). **Moderate risk**: hazard ratios and confidence intervals must be extrapolated from survival curves or point-estimates. **Low risk**: hazard ratios with confidence intervals are reported or obtained from individual patient data.

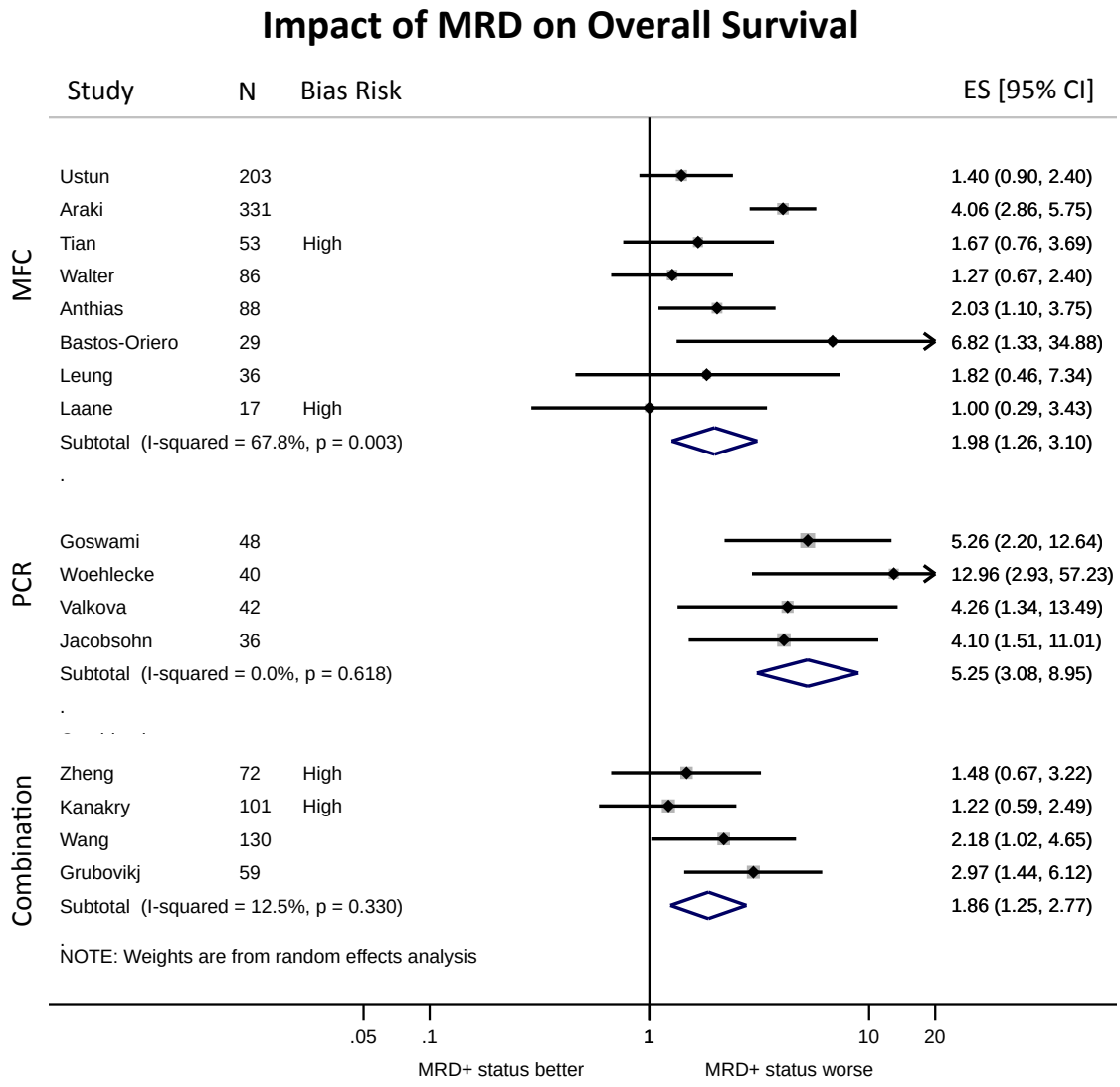
Supplementary Table 3: Details of conditioning regimens, stem cell sources, and GVHD prophylaxis for each study

Study	Conditioning	Stem Cell Source	GVHD Prophylaxis
Bleyzac et al[26]	MA: TBI or Bu-based	43% MRD 14% MURD 43% MMURD	CsA + 7.5 mg/kg rabbit ATG for URD + corticosteroids for CBT
Ustun et al[27]	MA: Cy/TBI [1320 cGy] or Flu/Bu + melphalan or Bu/Cy. *or* RIC: Cy/Flu/TBI [200 cGy] or Flu/Bu ± ATG.	MRD or UCB	CsA + MMF or sirolimus + MMF for UCB or RIC. CsA + MTX for others.
Zheng et al[25]	MA: Bu [12.8 mg/kg] / Cy [120 mg/kg] + HDAC *or* MA: Cy [120 mg/kg] / TBI [1200 cGy] + HDAC	UCB (mostly single unit)	CsA + MMF
Araki et al[24]	MA: various – Bu/Cy ± low-dose TBI, Bu/Flu, Bu/etoposide, Bu/clofarabine, high-dose TBI ± Cy or Flu, high-dose TBI/thiotepa/Flu, treosulfan/Flu ± low-dose TBI, Flu/low-dose TBI + radiolabeled antibody ± Cy	40% MRD 60% unrelated donor	CI + MTX (73%), CI + MMF (11%), Cy ± CI ± MMF (13%), other (3%)
Goswami et al[31]	MA (92%): Cy [120 mg/kg] / TBI or Flu [125 mg/m^2] / Cy / TBI [1200-1360 cGy except 600 cGy for older adults in some cases] *or* RIC: Flu-based [125 mg/m^2]	Mostly MRD, T-cell depleted	CsA
Rossi et al[19]	MA: Cy [120 mg/kg] / TBI [1200 cGy] *or* MA: Bu [9.6 mg/kg] + tepadine [10 mg/kg] + Flu [150 mg/m^2]	54% MRD, 33% MURD, 13% MMRD	CI + MTX + ATG. Cy added for MMRD
Tian et al[28]	MA: Bu [9.6 mg/kg] / Cy [3.6 g/m^2] *or* MA: TBI [750 cGy] / Cy [3.6 g/m^2] *or* MA: Bu or TBI/Cy + rabbit ATG	53% MRD 21% MURD 26% MMRD Some had planned DLI day 26 provided no GVHD	CSA + MMF + MTX
Walter et al[29]	RIC: low-dose TBI ± Flu or clofarabine	44% MRD 66% unrelated donor	CI + MMF ± rapamycin
Woehlecke et al[30]	MA (subset): mostly TBI or Bu-based	30% MRD 46% MURD 24% MMURD	Not listed
Anthias et al[22]	MA: Cy / TBI ± alemtuzumab (for URD) *or* RIC Flu + melphalan + alemtuzumab	34% MRD 57% unrelated donor 9% UCB	Not listed
Bastos-Oriero et al[33]	Not listed (86% MA)	46% MRD 32% MURD 15% UCB 7% MMRD	Not listed
Kanakry et al[21]	MA: Bu [targeted] /Cy [100 mg/kg] or Bu [targeted] / Flu [160 mg/m^2]	57% MRD 43% MURD	Post-transplant Cy [50 mg/kg days +3 and +4]

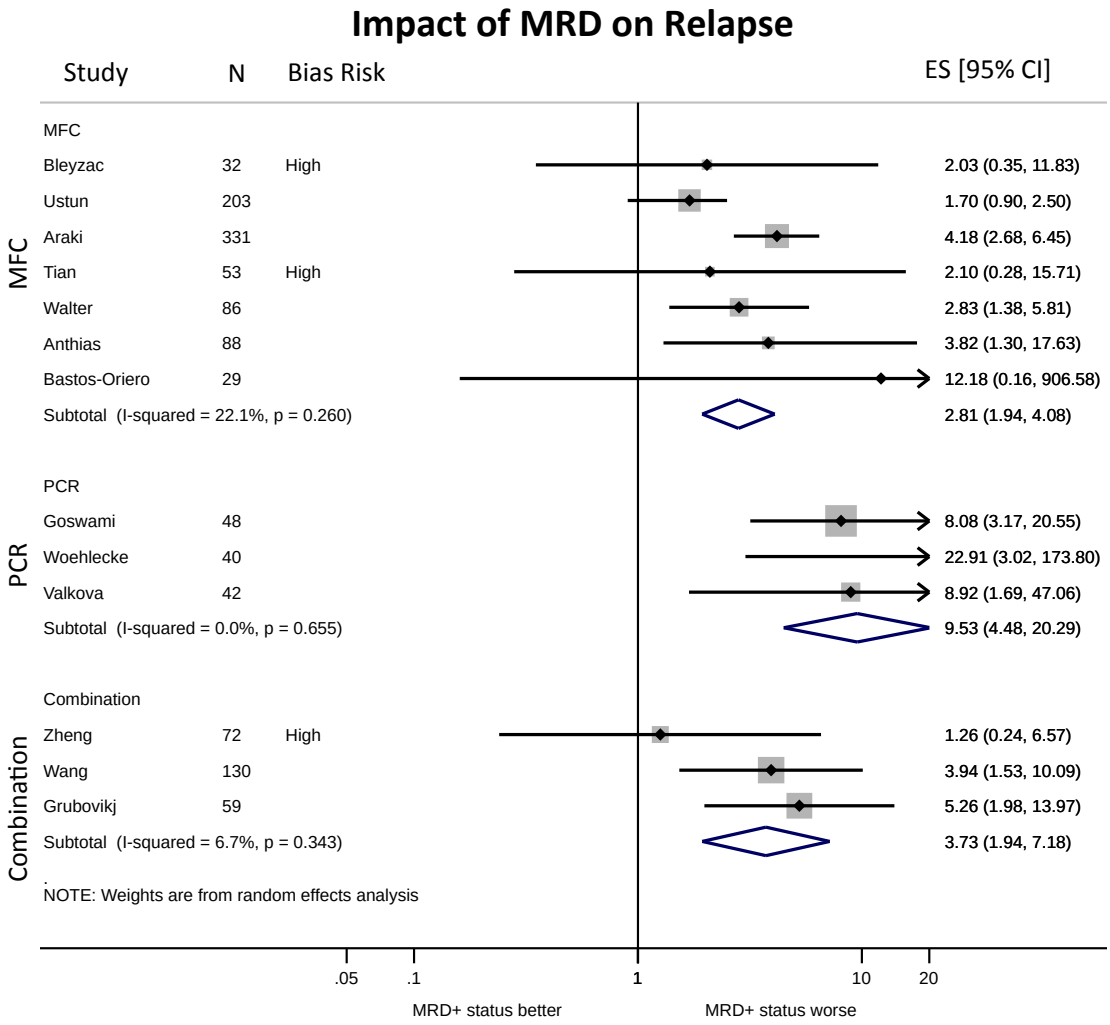
Wang et al[34]	MA: cytarabine [8 g/m^2] + Bu [9.6 or 12 mg/kg] + Cy [3.6 g/m^2] + semustine [250 mg/m^2] ± ATG	100% MMRD	CSA + MMF + MTX
Grubovikj et al[35]	88% MA 46% TBI-based	57.6% related 78% matched	Not listed
Leung et al[36]	MA: TBI/Cy (for matched) *or* MA: TBI-based or Flu + melphalan-based regimens (for haploidentical)	Not listed T-cell depletion used for haploidentical	CsA + MTX or MMF
Valkova et al[32]	MA: TBI or Bu-based *or* RIC: Flu + Bu or melphalan or TBI [200 cGy]	38% MRD 38% MURD 24% MMURD	CI ± MMF
Candoni et al[17]	RIC: Flu/Bu, Cy/thiotepa, treosulfan/Flu	56% MRD 38% URD 6% UCB	Not listed
Jacobsohn et al[18]	MA: TBI [1200 cGy] / Cy [120 mg/kg] + etoposide [1 g/m^2] *or* MA: Bu [12.8 mg/kg] / Cy [240 mg/kg] *or* RIC: Flu [180 mg/m^2] / Bu [targeted] + rabbit ATG	36% MRD 17% MURD 47% UCB	Not listed
Laane et al[37]	Not listed (100% MA)	Not listed	Not listed

Abbreviations: ATG, anti-thymocyte globulin; Bu, busulfan; CBT, cord blood transplant; CI, calcineurin inhibitor; CsA, cyclosporine A; Cy, cyclophosphamide; Flu, fludarabine; GVHD, graft-versus-host disease; HDAC, high-dose cytarabine; MA, myeloablative; MMF, micophenolate mofetil; MMRD, mismatched related donor; MMURD, mismatched unrelated donor; MRD, matched related donor; MTX, methotrexate; MURD, matched unrelated donor; RIC, reduced intensity conditioning; TBI, total body irradiation

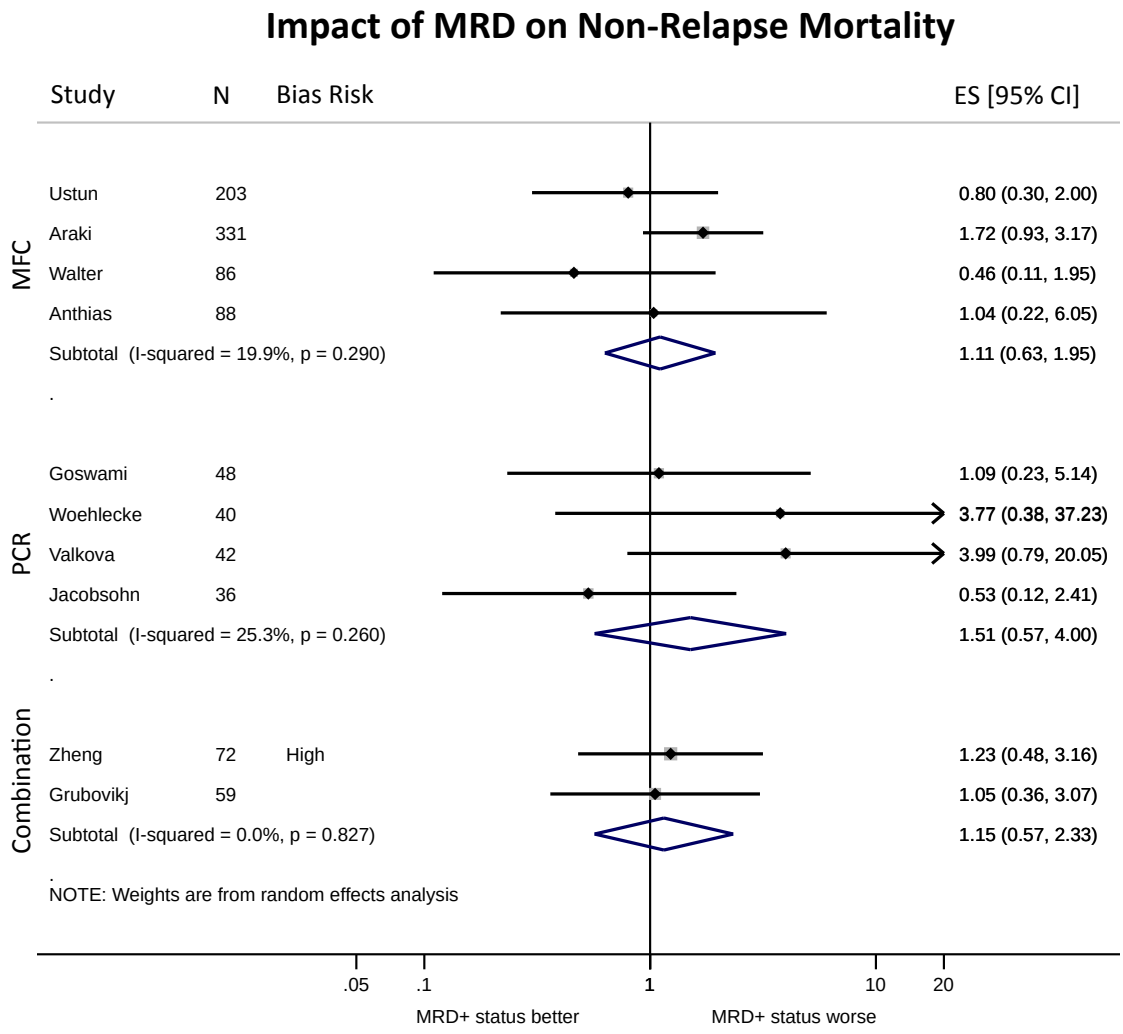
Supplemental Figure 1: Forest plot showing hazard ratio (effect size, ES) for overall survival with pooling of results for each MRD detection method. Columns indicate study size (N) and whether each study carries a high risk of bias (Bias Risk). Within each section, studies are listed by year of publication.



Supplemental Figure 2: Forest plot showing hazard ratio (effect size, ES) for cumulative incidence of relapse with pooling of results for each MRD detection method. Columns indicate study size (N) and whether each study carries a high risk of bias (Bias Risk). Within each section, studies are listed by year of publication.



Supplemental Figure 3: Forest plot showing hazard ratio (effect size, ES) for non-relapse mortality with pooling of results for each MRD detection method. Columns indicate study size (N) and whether each study carries a high risk of bias (Bias Risk). Within each section, studies are listed by year of publication.



Supplemental Figure 4: Forest plot showing hazard ratio (effect size, ES) for leukemia-free survival with pooling of results for studies using predominantly myeloablative and exclusively non-myeloablative conditioning strategies. Columns indicate year of publication (Year), study size (N), and method of MRD detection (Method).