Supplemental Digital Content* | Methodology | February 2017

Initial Diagnostic Workup of Acute Leukemia

Guideline from the College of American Pathologists and the American Society of Hematology

Corresponding Author:
Daniel A. Arber, MD

Authors:
James W. Vardiman, MD
R. Bryan Rumble, MSc
Anthony T. Smith, MLS
Nicole E. Thomas, MPH, CT(ASCP)[cm]

http://www.archivesofpathology.org/doi/pdf/10.5858/arpa.2016-0504-CP

*The Supplemental Digital Content was not copyedited by *Archives of Pathology & Laboratory Medicine.*

METHODS USED TO PRODUCE THE GUIDELINE

Panel Composition

The College of American Pathologists (CAP) Pathology and Laboratory Quality Center (the Center) and the American Society of Hematology (ASH) convened an expert panel (EP) consisting of general pathologists, board-certified hematopathologists, hematologists, a hematologist/oncologist, and a methodologist consultant to develop an evidence-based guideline to address the initial workup of acute leukemia. CAP and ASH approved the appointment of the project chair and panel members. The EP members performed the systematic evidence review.

An advisory panel (AP) of one patient advocate, one cytogeneticist, three hematologists/oncologists (including one pediatric hematologist/oncologist), one medical oncologist, and two hematopathologists also helped in the development of the guideline. The role of the AP members was to provide guidance and feedback on the key questions for the literature search, vet the draft guideline statements prior to the public comment period, and to review and provide feedback for the manuscript and supplemental digital content (SDC).

Conflict of Interest (COI) Policy

Prior to acceptance on the expert or advisory panel, potential members completed the CAP conflict of interest (COI) disclosure process, whose policy and form (in effect April 2010) require disclosure of material financial interest in, or potential for benefit of significant value from, the guideline's development or its recommendations 12 months prior through the time of publication. The potential members completed the COI disclosure form, listing any relationship that could be interpreted as constituting an actual, potential, or apparent conflict. The CAP Center uses the following criteria:

Nominees who have the following conflicts may be excused from the panel:

a. Stock or equity interest in a commercial entity that would likely be affected by the guideline or white paper
b. Royalties or licensing fees from products that would likely be affected by the guideline or white paper
c. Employee of a commercial entity that would likely be affected by the guideline or white paper

Nominees who have the following potentially manageable direct conflicts may be appointed to the panel:

a. Patents for products covered by the guideline or white paper
b. Member of an advisory board of a commercial entity that would be affected by the guideline or white paper
c. Payments to cover costs of clinical trials, including travel expenses associated directly with the trial
d. Reimbursement from commercial entity for travel to scientific or educational meetings

Everyone was required to disclose conflicts prior to beginning and continuously throughout the project's timeline. Expert panel members' disclosed conflicts are listed in the appendix of the manuscript. The CAP and ASH provided funding for the administration of the project; no industry funds were used in the development of the guideline. All panel members volunteered their time and were not compensated for their involvement, except for the contracted methodologist.

Literature Review and Analysis

The EP met 23 times through teleconference webinars from June 8, 2011 through August 16, 2016. Additional work was completed via electronic mail. The panel met in person July 19, 2013 to review evidence to date and draft recommendations and again September 15, 2015 to review feedback from the public comment period and to finalize the recommendations.

Prior to the first in-person meeting, the expert panel formed the following key questions (KQ) for which to base the literature search:

1. What clinical and laboratory information should be available during the initial diagnostic evaluation of a patient with acute leukemia?
2. What specimens and sample types should be evaluated during the initial workup of a patient with acute leukemia?
3. At the time of diagnosis, what tests are required for all patients for the initial evaluation of an acute leukemia?
4. Which tests should be performed only on a subset of patients, including in response to results of initial tests and morphology?
5. Where should laboratory testing be performed?
6. How should test results and the diagnosis be correlated and reported?

All expert panelists participated in the systematic evidence review (SER): title-abstract screening, full-text review, and data extraction of high-level studies (i.e., randomized control trials, systematic reviews, and clinical practice guidelines). A dual review was performed for each study and in each phase of the SER; the co-chairs adjudicated all conflicts. After assessing the data, the panel determined that the high-level studies alone and the initial data extraction points were not sufficient to draft guideline statements. The panel then decided to develop additional data extraction points to be included in the systematic review form and to include lower-level study designs. CAP staff and two contracted methodologists performed a second round of data extraction for the high-level studies and 150 cohort studies. The expert panel co-chairs performed an audit to verify accuracy and completeness. A total of 119 studies comprised the final body of studies included in the SER. Supplemental Figure 1 displays the results of the literature review. All articles were available as discussion or background references. All members of the EP participated in developing draft recommendations, reviewing open comment feedback, finalizing and approving the final recommendations, and writing/editing of the manuscript.

Peer Review

An open comment period was held from August 10 through August 31, 2011 on the ASH Web site www.hematology.org. Twenty-nine draft recommendations, four demographic questions, and three questions to assess feasibility were posted for peer review. An announcement was sent to the following societies deemed to have interest:

- American Association for Clinical Chemistry (AACC)
- American Cancer Society
- American College of Medical Genetics and Genomics (ACMG)
- American Society of Hematology (ASH)
- American Society for Clinical Oncology (ASCO)
- American Society for Clinical Pathology (ASCP)
- Association of Community Cancer Centers (ACCC)
- Association for Molecular Pathology (AMP)
- Association of Pathology Chairs (APC)
- Association of Directors of Anatomic and Surgical Pathology (ADASP)
- Cancer and Leukemia Group B (CALGB)
- Cancer Leadership Council
- Cancer Research and Prevention Foundation
- Canadian Association for Pathology (CAP-ACP)

- Canadian Partnership Against Cancer
- CAP Board of Governors, Councils, Committees and Membership
- Centers for Medicare & Medicaid Services (CMS)
- Center for Strategic Scientific Initiatives (CSSI)
- Children's Oncology Group (COG)
- Clinical Laboratory Improvement Advisory Committee (CLIAC) (CDC)
- Clinical Laboratory Management Association (CLMA)
- Division of Cancer Treatment and Diagnosis
- Division of Laboratory Programs, Standards, and Services
- Eastern Cooperative Oncology Group (ECOG)
- European Society for Medical Oncology (ESMO)
- European LeukemiaNet (ELN)
- International Society for Laboratory Hematology
- Leukemia and Lymphoma Society
- National Comprehensive Cancer Network (NCCN)
- National Health Council
- Office of Cancer Centers (OCC)
- Society for Hematopathology/European Association of Haematopathology (SH-EAHP)
- Society to Improve Diagnosis in Medicine
- Southwest Oncology Group (SWOG)
- Union for International Cancer Control (UICC)
- United States & Canadian Academy of Pathology (USCAP)
- US Food and Drug Administration (FDA)

"Agree" and "Disagree" responses were captured for every proposed recommendation. The website also received 789 written comments. Twenty-six draft recommendations achieved more than 90% agreement, two draft statements achieved more than 80% agreement, and one received more than 70% achievement. Each EP member was assigned three draft statements for which they had to review the comments and present them to the entire panel for group discussion. After consideration of the comments, two draft recommendations were maintained with the original language, 25 were revised, and two draft recommendations were combined into other statements which resulted in a total of 27 final recommendations. Resolution of all changes was obtained by majority consensus of the panel using nominal group technique (discussion at an in-person meeting, rounds of teleconference webinars, email discussion and multiple edited recommendations) amongst the panel members. The final recommendations were approved by the expert panel with a formal vote. The panel considered laboratory efficiency and feasibility throughout the entire considered judgment process.[1] 52% (89 of 171) responded that all of the draft guideline was feasible, 48% (82 of 171) responded that parts of it were feasible, and 0% responded that none of it was feasible. Neither formal cost analysis nor cost effectiveness models were performed.

An independent review panel (IRP) was assembled to review and approve the guideline for the CAP. The IRP was masked to the expert panel and vetted through the COI process. The ASH Executive Committee approved the guideline on behalf of the ASH.

Dissemination Plans

CAP plans to host an Initial Diagnostic Workup of Acute Leukemia resource page which will include a link to the manuscript and supplement; a summary of the recommendations, a teaching PowerPoint (Microsoft Corporation, Redmond, WA), a frequently asked question (FAQ) document, and an infographic. The guideline will be promoted and presented at various society meetings.

Systematic Evidence Review (SER)

The objective of the SER was to identify articles that provided data to inform the recommended testing for the proper diagnosis and optimal prognosis of acute leukemia, particularly pediatric and adult cases of acute myeloid leukemia (AML), acute lymphocytic leukemia (ALL), and mixed phenotype acute leukemia (MPAL). If of sufficient quality, findings from this review would provide an evidence-base to support the recommendations of the guideline. The scope of the SER and the key questions (KQs) were established by the EP in consultation with the methodologist prior to beginning the literature search.

Search and Selection

A systematic literature search was completed on 10/4/2011 for relevant evidence utilizing OvidSP, PubMed, and Science Direct to identify literature published from January 2005 through September 2011. The search strategy included medical subject headings (MeSH) and text words to capture the general concepts of the workup, diagnosis, prognosis, and management of acute leukemia (AL). The literature search was limited to the English language and human subjects. Publication type limits were added to capture clinical trials, meta-analyses, multicenter studies, cohort studies, prospective studies, and systematic reviews, and to exclude case reports. The electronic database searches were supplemented by a search for grey literature utilizing the TRIP database, the Cochrane Library, Guidelines International Network, and National Guideline Clearinghouse; expert panel recommendations; and a review of reference lists of included articles. A refresh of the OvidSP search strategy was completed on 4/24/2013 to identify relevant literature published from September 2011 to April 2013 and again on 08/11//2015 to identify relevant literature published from April 2013 to August 2015.The Ovid search strategy is included as Supplemental Figure 2. The PubMed and Science Direct searches were adaptations of the OvidSP strategy.

Selection at all levels was based on predetermined inclusion/exclusion criteria.
Included:

- Human studies
- Studies published in English

Not included:

- Focus of study was beyond scope of project even though the study group is patients with acute leukemia
- Focus of the study was response to a specific drug or drug combination, including phase I and II studies
- Focus of the study was side effects of leukemia therapies
- Focus of the study was on second malignancies after treatment of acute leukemia
- Focus of the study was on non-acute leukemia, including adult T-cell leukemia/lymphoma, large granular lymphocytic leukemia or B-lymphocytosis
- Focus of the study was infections or treatment of infections associated with acute leukemia
- Focus of the study was on myelodysplastic syndromes (MDS)
- Focus of the study was leukemia in a specific nationality, city or state
- Focus of the study was a specific treatment protocol (including hematopoietic stem cell transplant), treatment complication, or “novel therapy”
- Focus of the study was a specific organ function, other than bone marrow, peripheral blood, csf, or spleen
- Focus of the study was on relapsed acute leukemia
- Focus of the study was on the methodologic aspects of a laboratory test, rather than on the significance of test results
- The study was a cell line study with no clinical findings
- The study focused on predisposition for acute leukemia
- Focus of study was on lymphoma or other malignancy
- Focus of the study was on the mechanisms of leukemogenesis

- Non-English-language article/document or an English-language abstract or summary without a full article/document available in English.
- Studies published prior to 2005
- Studies that were Veterinary/Non-human
- Studies that were case reports
- Studies that had technical or study design concerns
- Cohort studies with fewer than 100 patients
- Clinical studies where a protocol was not referenced

Outcomes of Interest

For KQ1 the outcomes of interest included:

- survival rates (e.g., overall survival [OS], disease free survival [DFS]) expressed in "> x" years when evaluating survival in pediatric populations.

For KQ2 the outcomes of interest included:

- utility and technical requirements of bone marrow aspirate for diagnosis of AL
- utility and technical requirement of core biopsy for diagnosis of AL
- utility of bone marrow clot section for the diagnosis of AL
- utility of bone marrow touch preparation for the diagnosis of AL

For KQ3 the outcomes of interest included:

- utility of antigens for the diagnosis of AML, APL, and ALL
- utility of minimal residual disease (MRD) in AML, ALL, and MPAL
- significant differences in blood versus marrow for flow cytometry in diagnosis of AL
- significant differences in blood versus marrow for MRD
- utility of antigens in detection of MRD in AML, ALL, and MPAL
- differences in MRD by flow cytometry versus MRD by molecular studies/sequences
- utility of engraftment studies for detection of MRD after transplant for AL
- antigens detected by flow cytometry for therapeutic target in AML, ALL, and MPAL

For KQ4 the outcomes interest included:

- survival rates by test type (MRD, fluorescent in situ hybridization [FISH], polymerase chain reaction [PCR], quantitative PCR, genetic and molecular testing, gene expression analysis, mutational analysis, immunohistochemistry, flow cytometry immunophenotyping, karyotyping, and various translocations)

For KQ5 and KQ6 the outcomes of interest included non-traditional outcomes. The outcomes for these included:

- differences in diagnosis or in test results when duplicate tests were performed in more than one institution (i.e., primary care institution versus treatment center)
- classification scheme for reporting AL

Data Extraction & Management

The data elements from an included article/document were extracted by one reviewer into standard data formats and tables developed using the systematic review database software, DistillerSR (Evidence Partners Inc., Ottawa, Canada); a second reviewer confirmed accuracy and completeness. Any discrepancies in data extraction were resolved by discussion between the co-chairs and the methodologist. A bibliographic database was established in EndNote (Thomson Reuters, Carlsbad, CA) to track all literature identified and reviewed during the study.

While most of evidence tables are included in the manuscript, five appear in this SDC. Supplemental Table 1 and 2 inform Statement 1 and have only four studies each that inform the collection of patient history and ethnicity and performance status. Supplemental Table 3 contains only four studies and informs Statement 2. Supplemental Table 4 informs Statement 20. Supplemental Table 5, which contains no p-values, informs Statement 23.

Quality Assessment Methods

An assessment of the quality of the evidence was planned for all retained studies following application of the inclusion and exclusion criteria. Using this method, studies deemed be of low quality would not be excluded from the systematic review, but would be retained and their methodological strengths and weaknesses discussed where relevant. Studies would be assessed by confirming the presence of items related to both internal and external validity, and which are all associated with methodological rigor and a decrease in the risk of bias. These items were assessed as being either yes, no, partial, not reported (NR), or not applicable (N/A) in the following way:

Clinical practice guidelines (CPGs) were assessed for quality by confirming the following attributes were considered and incorporated in its design as recommended by the Institute of Medicine (IOM).[2]

- Based on a systematic review
- Included a multidisciplinary panel
- Patient preferences were considered
- Important patient sub-types were considered
- Methods were well-described and reproducible
- Information on potential conflicts of interest were gathered and disclosed
- Quality of the evidence was assessed
- Strength of the evidence was rated
- CPG includes a plan for updating
- Sources of funding are disclosed

Meta-analyses were assessed in a similar fashion to CPGs according to the following criteria:

- Based on a systematic review
- Methods were well-described and reproducible
- Quality of the evidence was assessed
- Any planned pooling was stated a priori
- Limitations of the analysis are discussed
- Sources of funding are disclosed

Randomized controlled trials (RCTs) and Quasi-RCTs were assessed for quality according to reporting and full description of:

- Randomization method fully-described
- Details on any blinding were provided
- Provided details of all planned analyses
- Stated the expected effect size and described the statistical power calculation
- Reported the length of follow-up
- Provided a description of the baseline characteristics for all patients by treatment/assessment arm
- Sources of funding are disclosed

Non-randomized clinical trials (NRCTs) and prospective cohort studies (PCSs) were assessed according to:

- Balance between treatment/assessment groups
- Reporting of baseline characteristics
- Reporting if any adjustments were made where baseline differences were detected
- Sources of funding

Each study was assessed individually, and then each study type was summarized. Finally, a summary of the overall quality of the evidence was given considering the evidence in totality.

Supplemental Table 6 summarizes the quality assessment criteria by study design.

Quality Assessment Results

A total of 119 studies[3-121] were included in our systematic review. This body of evidence comprised two meta-analyses,[37, 51] four RCTs,[49, 62, 109, 112] six NRCTs,[18, 46, 47, 115, 116, 119] and 107 PCS.[3-17, 19-36, 38-45, 48, 50, 52-61, 63-108, 110, 111, 113, 114, 117, 118, 120, 121] In the following sections, the quantity of the evidence as determined by the number of studies that met our inclusion criteria and were retained, the evidence type as determined by study design, the quality of that evidence as determined by the risk of bias assessment, and its consistency are all reported, both as individual studies and in totality, statement by statement.

Statement 1 is supported by a total of 30 studies, comprising two NRCT[115, 116] and 28 PCS[5, 11, 14, 15, 21-23, 25, 26, 28, 30, 36, 42, 45, 48, 50, 55, 58, 59, 70, 73, 74, 85, 94, 101, 105, 114, 118] inform collecting the patient's age. The two NRCTs[115, 116] were both Single cohort studies. Neither reported on the baseline characteristics of the patient population. Both received Non-industry funding, and both were deemed to have a Low-moderate risk of bias. For the 28 PCS,[5, 11, 14, 15, 21-23, 25, 26, 28, 30, 36, 42, 45, 48, 50, 55, 58, 59, 70, 73, 74, 85, 94, 101, 105, 114, 118] 20[14, 21, 23, 26, 30, 36, 42, 45, 48, 50, 55, 58, 73, 74, 85, 94, 101, 105, 114, 118] were Single cohort designs, and eight[5, 11, 15, 22, 25, 28, 59, 70] used two or more comparison groups. All of the PCS reported on baseline characteristics, except for the studies reported by Moorman et al[114] and Lo-Coco et al.[42] Only two[5, 70] reported that adjustments were made when differences in baseline characteristics were detected. All of the PCS reported Non-industry sources of funding, except for the studies by Schneider et al[118] and Harvey et al[85] which both reported at least partial industry funding. Five studies[14, 36, 59, 74, 94] did not report the source of funding, if any. Most of the PCS had a risk of bias determination of Low-moderate, except for three studies determined to be low[5, 45, 70] but seven[14, 36, 50, 59, 74, 85, 94] were determined to be moderate. None of these studies were found to have methodological flaws that would raise concerns about the studies' findings.

Refer to Supplemental Table 7 for the quality assessment (QA) results for studies informing Statement 1.

Statement 2 is supported by four[26-28, 36] PCS. Two[26, 36] were single-cohort studies allowing for within-group comparisons, and two[27, 28] used two or more groups allowing for between-group comparisons. Neither of the two studies that compared between groups reported balance between comparison groups, but the study by Schmiegelow K et al, 2010 reported making adjustments to account for these differences.[27] All of the studies reported on baseline patient characteristics. Three[26-28] reported Non-industry sources of funding, although one[36] did not disclose the source of funding, if any. One[27] study was deemed to have a low risk of bias, two[26, 28] were deemed to have a Low-moderate, and one[36] was deemed to have a moderate risk of bias. None of these studies were found to have methodological flaws that would raise concerns about the studies' findings.

Refer to Supplemental Table 8 for the QA results for studies informing Statement 2.

Statement 3 is supported by 51 studies,[3-8, 10, 11, 14, 17, 18, 21-23, 25-28, 30, 32, 35, 38, 40, 42, 45-47, 49, 50, 53, 59-65, 67, 70, 73, 89, 90, 97, 100, 102, 103, 107, 110, 116, 117, 119] comprising two RCT,[49, 62] five NRCT,[18, 46, 47, 116, 119] and 44 PCS.[3-8, 10, 11, 14, 17, 21-23, 25-28, 30, 32, 35, 38, 40, 42, 45, 50, 53, 59-61, 63-65, 67, 70, 73, 89, 90, 97, 100, 102, 103, 107, 110, 117]
For the two RCT,[49, 62] neither provided details on the randomization, other than to confirm the studies were randomized. Only one[49] provided any details on planned analyses, expected effect size, power and sample size calculation, and length of follow-up. Both did report on differences in patient characteristics. Both reported Non-industry sources for funding. The trial by Lange et al[49] was deemed to have a very low risk of bias and the trial by Schneider et al[62] was deemed to have a high risk of bias.

For the five NRCT,[18, 46, 47, 116, 119] two[116, 119] were Single cohort designs, allowing for within group comparisons, and two[46, 47] used two or more groups, allowing for between group comparisons. Neither of these two reported a balance between arms, and neither made adjustments based on these differences. Only two[46, 47] reported on baseline characteristics. All reported Non-industry sources of funding. Four[46, 47, 116, 119] reported a risk of bias of Low-moderate and one[18] reported a moderate risk of bias.

For the 44 PCS,[3-8, 10, 11, 14, 17, 21-23, 25-28, 30, 32, 35, 38, 40, 42, 45, 50, 53, 59-61, 63-65, 67, 70, 73, 89, 90, 97, 100, 102, 103, 107, 110, 117] 29 were Single cohort designs,[3, 14, 17, 21, 23, 26, 30, 32, 35, 38, 40, 42, 45, 50, 53, 60, 61, 63, 65, 67, 73, 89, 97, 100, 102, 103, 107, 110, 117] allowing for within group comparisons, and 15 used two or more groups,[4-8, 10, 11, 22, 25, 27, 28, 59, 64, 70, 90] allowing for between group comparisons. Of these 15, only two[64, 70] reported balanced groups. Of the 13 remaining, only five[4-6, 27, 70] reported making adjustment in any analyses to account for these differences. All of the studies reported Non-industry funding, except for four[14, 59, 100, 110] that did not report the source of funding, if any. Seven of the studies were deemed to have a low risk of bias,[4-6, 27, 45, 64, 70] 30 were deemed to have a Low-moderate risk of bias,[3, 7, 8, 10, 11, 17, 21-23, 25, 26, 28, 30, 32, 35, 38, 40, 42, 53, 60, 61, 63, 65, 67, 73, 89, 97, 102, 103, 107] and seven were deemed to have a moderate risk of bias.[14, 18, 50, 59, 90, 100, 110] None of these studies were found to have methodological flaws that would raise concerns about the studies' findings.

Refer to Supplemental Table 9 for the QA results for studies informing Statement 3.

Statement 4 is axiomatic. Bone marrow aspirates are among the most common specimens used in the diagnosis of AL and most studies included in data extraction used a bone marrow aspirate. Therefore, no quality assessment was performed.

No evidence from the SR informed Statement 5 or 6.

Statement 7 is supported by four PCS[75, 76, 87, 96] that met the inclusion criteria for our systematic review. Two of these were Single cohort designs allowing for within group comparisons,[87, 96] and two compared between two groups.[75, 76] Of the two studies with two groups, one[75] reported balanced groups and although the other[76] was imbalanced, it did make adjustments to its analysis to account for these differences. All of the studies reported on baseline characteristics, and all reported Non-industry funding sources. Two[75, 76] of the studies were deemed to have a low and the other two[87, 96] were deemed to have a Low-moderate risk of bias. None of these studies were found to have methodological flaws that would raise concerns about the studies' findings.

Refer to Supplemental Table 10 for the QA results for studies informing Statement 7.

No evidence from the SR informed Statement 8, 9, 10, or 11.

Statement 12 is supported by 15 studies that met the inclusion criteria for our systematic review[12, 15, 25, 27-29, 31, 33, 36, 52, 88, 93, 98, 106, 112] comprising one RCT[112] and 14 PCS.[12, 15, 25, 27-29, 31, 33, 36, 52, 88, 93, 98, 106]

The single RCT, reported by Yin et al, 2012 did not provide details on the randomization other than noting that it was a randomized trial, blinding (if any), nor the expected effect size and the power calculation.[112] It did report on all the planned analyses, length of follow-up, and differences in baseline patient characteristics. The funding source was reported to be Non-industry. Overall, this trial was deemed to have a moderate risk of bias.

For the 14 PCS,[12, 15, 25, 27-29, 31, 33, 36, 52, 88, 93, 98, 106] nine[12, 29, 31, 33, 36, 52, 88, 93, 106] reported on single-cohort designs that allowed for within-group comparisons only, and five[15, 25, 27, 28, 98] reported on studies that contained two or more groups which allowed for between-group comparisons. Only one[98] of these studies that used two or more groups reported a balance between groups, and of

the other four[15, 25, 27, 28] only one[27] reported making any adjustments to the analyses to account for these differences. Twelve[12, 15, 25, 27-29, 31, 33, 36, 52, 93, 98] reported on baseline characteristics. Eleven[15, 25, 27-29, 31, 33, 52, 88, 93, 106] reported Non-industry funding, one[98] reported at least partial industry funding, and two[12, 36] did not disclose the source of funding, if any. Two[27, 52] of these studies was deemed to have a low risk of bias, ten[15, 25, 28, 29, 31, 33, 88, 93, 98, 106] were deemed to have a low moderate, and two[12, 36] were deemed to have a moderate risk of bias.

Overall, none of these 15 studies were found to have methodological flaws that would raise concerns about the studies' findings. Refer to Supplemental Table 11 for the QA results for studies informing Statement 12.

Statement 13 is supported by six PCS[25-28, 44, 97] that met the inclusion criteria for our systematic review. Three of these studies[26, 44, 97] used a Single cohort that provided within group comparisons, and the other three[25, 27, 28] used more than one cohort allowing for between group comparisons. The three that used between group comparisons all reported imbalances between groups, but only one, the study reported by Schmiegelow et al[27] reported adjustments to any analyses based on these differences. All of the studies reported on the baseline characteristics of the study population. None of the studies reported any industry funding. Risk of bias assessments ranged from low[27, 44] to Low-moderate.[25, 26, 28, 97] None of these studies were found to have methodological flaws that would raise concerns about the studies' findings. Refer to Supplemental Table 12 for the QA results for studies informing Statement 13.

Statement 14 is supported by two PCS[23, 108] that met the inclusion criteria for our systematic review. Both of these studies used a Single cohort, allowing for within group comparisons, and therefore no adjustments were needed to account for differences between comparative groups. Both of these studies reported on baseline characteristics. Both of these studies reported Non-industry sources of funding, and were deemed to have a Low-moderate risk of bias. None of these studies were found to have methodological flaws that would raise concerns about the studies' findings. Refer to Supplemental Table 13 for the QA results for studies informing Statement 14.

Statement 15 is supported by 14 prospective cohort studies.[9, 30, 34, 56, 69, 77, 79, 84-86, 92, 98, 114, 122] Of these 14 studies, nine[9, 30, 34, 56, 77, 79, 85, 86, 114] were Single cohort designs allowing for within-group comparisons. Five[69, 84, 92, 98, 122] compared two or more groups, and of these five, four[69, 84, 92, 122] reported imbalances between the comparison groups. Of these four, only one[92] made adjustments based on the detected imbalance. All but four[69, 84, 92, 114] of the 14 reported on patient baseline characteristics. The reported source of funding for all 14 studies was Non-industry, except for the studies reported by Mullighan et al,[98] 2009 and Harvey et al,[85] 2010, which were both at least partially industry supported. One[92] of the studies was deemed to have a low risk of bias, ten[9, 30, 34, 56, 77, 79, 86, 98, 114, 122] were deemed to have a Low-moderate, and three[69, 84, 85] were deemed to have a moderate risk of bias. None of these studies were found to have methodological flaws that would raise concerns about the studies' findings. Refer to Supplemental Table 14 for the QA results for studies informing Statement 15.

For Statement 16, *FLT3-ITD* testing is supported by 13 PCS[3, 8, 12, 14, 17, 20, 22, 24, 33, 71, 89, 91, 121] that met the inclusion criteria for our systematic review. Ten[3, 12, 14, 17, 20, 24, 33, 71, 89, 91] of the 13 studies used a Single cohort design allowing for within group comparisons, while the remaining three[8, 22, 121] used two or more groups, allowing for between group comparisons. Of the three studies that used two or more groups, two[8, 22] had an imbalance between the comparison groups, and neither reported any adjustments to any analysis to account for these differences. All 13 of the studies reported on baseline characteristics. All but two[12, 14] studies reported Non-industry funding, and these two did not disclose the source of funding, if any. Of the 13 studies, one[121] was deemed to have a low risk of bias, ten[3, 8, 17, 20, 22, 24, 33, 71, 89, 91] were deemed to have a Low-moderate, and two[12, 14] were deemed to have a moderate risk of bias. None of these studies were found to have methodological flaws that would raise concerns about the studies' findings. Refer to Supplemental Table 15 for the QA results for studies informing *FLT3-ITD* testing.

The recommendation for testing of other mutations in AML is supported by 21 studies,[4, 13, 20, 21, 38, 40, 59, 61, 63-65, 70, 71, 87, 115-117, 123-126] comprising one systematic review based meta-analysis,[125] three NRCT,[115, 116, 126] and 17 PCS .[4, 13, 20, 21, 38, 40, 59, 61, 63-65, 70, 71, 87, 117, 123, 124] The meta-analysis, reported by Zhou et al,[125] was based on a systematic review, with a well-reported methods section, and all pooling was planned and described a priori. It included an assessment of the quality of all the pooled studies, and discussed the limitations of the research. The funding source was Non-industry. This paper was deemed to have a low risk of bias.

The three NRCT[115, 116, 126] were all Single cohort designs allowing for within group comparisons only. For this reason, comparisons were made between equal groups, therefore no adjustments were required to account for differences. None of these reported on baseline characteristics, and all reported Non-industry sources of funding. All were deemed to have a Low-moderate risk of bias.

For the 17 PCS[4, 13, 20, 21, 38, 40, 59, 61, 63-65, 70, 71, 87, 117, 123, 124], 13 were Single cohort studies that allowed for within group comparisons only,[13, 20, 21, 38, 40, 61, 63, 65, 71, 87, 117, 123, 124] and four used two groups allowing for between group comparisons.[4, 59, 64, 70] Of these studies with two groups, two were balanced between arms,[64, 70] and two were not.[4, 59] Of the two that were not balanced between groups, the study reported by Metzeler et al[4] reported making adjustments in the analysis to account for these differences, while the study reported by Damm et al[59] did not. All of the PCS cohort studies reported on baseline characteristics. All reported Non-industry sources of funding, except for the study reported by Damm et al[59] which did not disclose the source of funding, if any. Of these studies, three were deemed to have a low risk of bias,[4, 64, 70] 13 were deemed to be Low-moderate,[13, 20, 21, 38, 40, 61, 63, 65, 71, 87, 117, 123, 124] and one was deemed to have a moderate risk of bias.[59]
None of these studies were found to have methodological flaws that would raise concerns about the studies' findings.

Refer to Supplemental Table 16 for the QA results for studies informing testing of other mutations in Statement 16.

Statement 17 is supported by two PCS[24, 33] obtained from the systematic review. Both of these were Single cohort designs reporting on within-group comparisons. Both of them reported on baseline patient characteristics, and both reported Non-industry funding. The risk of bias assessment for both was Low-moderate. Overall, none of the studies providing the evidence base for statement 17 were found to have methodological flaws that would raise concerns about the studies' findings. Refer to Supplemental Table 17 for the QA results for studies informing Statement 17.

No evidence from the SR informed Statement 18.

Statement 19 is supported by two PCS[3, 24] that met the inclusion criteria for our systematic review. Both of these studies used a Single cohort design allowing for within group comparisons, and for this reason, no adjustments for group differences were needed. Both studies reported on baseline characteristics. Both reported Non-industry funding. Both of these studies were deemed to have a Low-moderate risk of bias. Neither of the studies were found to have methodological flaws that would raise concerns about their findings. Refer to Supplemental Table 18 for the QA results for studies informing Statement 19.

Statement 20 is supported by 10 studies[4, 45, 68, 70, 75, 101, 103, 116, 119, 120], comprising two NRCTs[116, 119] and eight[4, 45, 68, 70, 75, 101, 103, 120] PCS. Both of the NRCTs[116, 119] were Single cohort design allowing for within group comparisons, and therefore, no adjustments were needed to account for between group differences. Neither of these studies reported on baseline characteristics, and both reported Non-industry funding. Both were deemed to have a Low-moderate risk of bias assessment.

Five of the studies were Single cohort designs,[45, 101, 103, 116, 119] allowing for within group comparisons. The remaining five studies used two groups allowing for between group comparisons.[4, 68, 70, 75, 120] Of these, two[70, 75] reported a balance between the groups and three[4, 68, 120] reported imbalances. Only the study reported by Metzeler et al[4] made adjustments to any analysis based on these differences. All of the prospective cohort studies reported Non-industry funding. Risk of bias assessments ranged from low[4, 45, 70, 75, 120] to Low-moderate.[68, 101, 103]

None of these studies were found to have methodological flaws that would raise concerns about the studies' findings. Refer to Supplemental Table 19 for the QA results for studies informing Statement 20.

No evidence from the SR informed Statement 21 or 22.

Statement 23 is supported by 28 studies,[4, 6, 8, 9, 14, 18, 19, 31, 36, 37, 42, 47, 48, 51, 54, 57, 60, 61, 71, 76, 80, 83, 94, 96, 108, 109, 111, 127] comprising two meta-analyses,[37, 51] one RCT,[109] two NRCT,[18, 47] and 23 PCS.[4, 6, 8, 9, 14, 19, 31, 36, 42, 48, 54, 57, 60, 61, 71, 76, 80, 83, 94, 96, 108, 111, 127] For the two meta-analyses,[37, 51] neither was based on a systematic review, but both had well-reported methods sections describing how the evidence was obtained. All pooling performed was planned a priori in both papers. Neither included an assessment of the quality of the studies that contributed to their pooled analyses, and neither discussed any of the limitations of their findings. Both reported Non-industry funding, and both were deemed to have a Low-moderate risk of bias.

One trial, a RCT reported by Vance et al[109] did not report on several key items such as details of the randomization (other than that it was a randomized trial), blinding, effect size and power calculations, length of follow-up, or differences in baseline patient characteristics. The funding source for this trial was reported to be Non-industry. Due to deficits in the reporting of this trial, it was deemed to have a moderate-high risk of bias.

For the two NRCT,[18, 47] the trial by Moorman et al[18] did not report on key items like the number of patients in each comparison group, baseline characteristics, and adjustments due to differences in either baseline characteristics or imbalances between comparison groups. This trial did report Non-industry funding, but due to the deficits in reporting, it was deemed to have a moderate risk of bias. The trial reported by Arico et al[47] did report the number of patients allocated to each arm, but the comparison groups were imbalanced, and no adjustments were made to account for this. Baseline characteristics were reported for each comparison group, and Non-industry funding supported this trial. It was deemed to have a Low-moderate risk of bias.

For the PCS, 17 were Single cohort designs,[9, 14, 19, 31, 36, 42, 48, 60, 61, 71, 80, 83, 94, 96, 108, 111, 127] which allowed for within group comparisons, and the other six included two or more groups allowing for between group comparisons.[4, 6, 8, 54, 57, 76] None of the studies comparing two or more groups were balanced between groups, but three reported making adjustments in the analyses to account for these differences.[4, 6, 76] All of the studies but six reported on baseline characteristics.[19, 42, 57, 60, 80, 111] All of the studies that disclosed funding sources reported Non-industry funding, although four[14, 36, 83, 94] did not report the source of funding, if any. Three[4, 6, 76] were deemed to have a low risk of bias, 14 were deemed to have a Low-moderate,[8, 9, 19, 31, 42, 48, 54, 60, 61, 71, 96, 108, 111, 127] and six[14, 36, 57, 80, 83, 94] were deemed to have a moderate risk of bias.

None of these studies were found to have methodological flaws that would raise concerns about the studies' findings. Refer to Supplemental Table 20 for the QA results for studies informing Statement 23.

Statement 24 is supported by two PCS[54, 57] that met the inclusion criteria for our systematic review. Both used two groups allowing for between group comparisons, but both reported imbalances between the comparison groups, and neither reported making adjustments to account for these imbalances. Only the study by Barbaric et al[54] reported on baseline

characteristics. Both reported Non-industry sources of funding. The study by Barbaric et al[54] was deemed to have a Low-moderate risk of bias and the study reported by Mrózek et al[57] was deemed to have moderate risk of bias. Neither of the studies was found to have methodological flaws that would raise concerns about their findings. Refer to Supplemental Table 21 for the QA results for studies informing Statement 24.

No evidence from the SR informed statements 25 or 26.

Statement 27 is supported by 40 PCS[3-5, 8, 9, 11-16, 19, 23, 24, 26, 28-30, 39, 66, 69, 72, 73, 79, 81, 82, 86, 89, 91, 95, 97, 99, 103, 104, 107, 108, 110, 111, 113, 114] obtained in the systematic review. The risk of bias assessment scores ranged from low[4, 5] to moderate.[12, 14, 66, 69, 81, 82, 110, 113] All but eight[4, 5, 8, 11, 15, 28, 69, 81] of the studies reported balance between the comparison groups, two of which reported that adjustments were made in any analyses to account for these differences.[4, 5] All but eight[19, 39, 69, 81, 82, 95, 111, 114] of the studies reported on baseline characteristics between groups. One of the studies reported at least partial industry funding,[113] although four[12, 14, 66, 110] did not report the source of funding, if any. Overall, none of the studies providing the evidence base for Statement 27 were found to have methodological flaws that would raise concerns about the studies' findings. Refer to Supplemental Table 22 for the QA results for studies informing Statement 27.

Overall, the body of evidence included in this CPG represents a methodologically rigorous and representative summary of the available evidence with an overall low to moderate risk of bias. Seventeen of the included studies were determined to have a low risk of bias,[4-6, 27, 43-45, 49, 52, 64, 70, 75, 76, 78, 92, 120, 121] 77 a Low-moderate risk of bias,[3, 7-11, 13, 15-17, 19-26, 28-35, 37-40, 42, 46-48, 51, 53-56, 58, 60, 61, 63, 65, 67, 68, 71-73, 77, 79, 86-89, 91, 93, 95-99, 101-108, 111, 114-119] 23 a moderate risk of bias,[12, 14, 18, 36, 41, 50, 57, 59, 66, 69, 74, 80-85, 90, 94, 100, 110, 112, 113] one a moderate-high risk of bias,[109] and one a high risk of bias.[62]

Assessing the Strength of Recommendations

The central question that the panel addressed in developing the guideline was "For the initial workup of acute leukemias, including AML, ALL, mixed phenotype acute leukemia, what is the recommended testing for proper diagnosis and prognosis determination?"

Development of recommendations required that the panel review the identified evidence and make a series of key judgments:

1) What are the significant findings related to each KQ or outcome? Determine any regulatory requirements and/or evidence that support a specific action.

2) What is the overall strength of evidence supporting each KQ or outcome? Strength of evidence is graded as Convincing, Adequate or Inadequate, based on four published criteria (Supplemental Table 23). Strength of evidence is a key element in determining the strength of a recommendation.

3) What is the strength of each recommendation? There are many methods for determining the strength of a recommendation based on the strength of evidence and the magnitude of net benefit or harm. However, such methods have rarely (if ever) been applied to the area of synoptic reporting. Therefore, the method for determining strength of recommendation has been modified for this application (Supplemental Table 24), and is based on the strength of evidence and the likelihood that further studies will change the conclusions. Recommendations not supported by evidence (i.e., evidence was missing or insufficient to permit a conclusion to be reached) were made based on consensus expert opinion. Another potential consideration is the likelihood that additional studies will be conducted that fill gaps in knowledge.

4) What is the net balance of benefits and harms? For each guideline statement, the panel considered the desirable effects, the undesirable effects, the resources required, and feasibility. Acceptability is addressed in the manuscript and in the discussion of the results of the public comment period for each guideline statement.

Discussion of Benefits and Risks of Implementing the Recommendations

Statement 1. The treating clinician should provide relevant clinical data or ensure that this is readily accessible by the pathologist.

Note: These data include, but is not limited to the patient's age, gender, ethnicity, history of any hematologic disorder or known predisposing conditions or syndromes, any prior malignancy, exposure to cytotoxic therapy, immunotherapy, radiotherapy, or other possibly toxic substances, and any additional clinical findings of diagnostic or prognostic importance. The treating clinician should also include any history of possibly confounding factors, such as recent growth factor therapy, transfusions or other medications that might obscure or mimic the features of acute leukemia. The treating clinician should also obtain and provide information regarding any family history of any hematologic disorder or other malignancies.

Providing relevant and clinical data helps to ensure that the pathologist has all of the information needed to inform the initial diagnosis of AL. The time it takes to collect and capture all of the proper information might be an undesirable effect, but the panel strongly believes that the benefits of having this information grossly outweigh any risks or harms. Most treating clinicians already have access to this clinical data and it is often found in the electronic medical record, and thus available to the pathologist as well. Therefore, this recommendation does not call for additional resources, is feasible, and acceptable to both treating clinicians and pathologists.

Statement 2. The treating clinician should provide relevant physical examination and imaging findings or ensure that these are readily accessible by the pathologist.

Note: This includes, but is not limited to, neurologic exam findings and the presence of tumor masses (e.g., mediastinal), other tissue lesions (e.g., cutaneous), and/or organomegaly.

The evidence shows that providing relevant data from clinical, physical examination, and imaging findings, aid in the initial diagnosis of acute leukemia and a more accurate diagnosis. The panel believes that there is little to no undesirable effects of providing information from these domains. Having the data accessible via the electronic medical record is optimal, but may require resources for some institutions. Currently, most clinicians have this information readily available, making the recommendation feasible.

Statement 3. The pathologist should review recent or concurrent complete blood counts (CBCs) and leukocyte differentials and evaluate a peripheral blood smear.

The results of the CBC and evaluation of the peripheral blood smear (PBS) aid in the diagnosis of AL and are necessary for predicting prognosis. The panel believes that there is little to no undesirable effects of reviewing/evaluating these samples. This is the standard of care and does not require any additional resources, making this recommendation feasible.

Statement 4. The treating clinician or pathologist should obtain a fresh bone marrow aspirate for all patients suspected of acute leukemia, a portion of which should be used to make bone marrow aspirate smears for morphologic evaluation. If performed, the pathologist should evaluate an adequate bone marrow trephine core biopsy, bone marrow trephine touch preparations, and/or marrow clots, in conjunction with the bone marrow aspirates.

Note: If bone marrow aspirate material is inadequate or if there is compelling clinical reason to avoid bone marrow examination, peripheral blood may be used for diagnosis and ancillary studies if sufficient numbers of blasts are present. If a bone marrow aspirate is unobtainable, touch imprint preparations of a core biopsy should be prepared and

evaluated, and an additional core biopsy may be submitted unfixed in tissue culture medium for disaggregation for flow and genetic studies. Optimally, the same physician should interpret the BM aspirate smears and the core biopsy specimens, or the interpretations of these specimens should be correlated if performed by different physicians.

The panel believes that recommending a fresh bone marrow aspirate for all patients suspected of AL will decrease diagnostic errors. An undesirable effect might be the risk of complications resulting from performing the bone marrow procedure, especially the core biopsy. The possibility of mis-diagnosis increases in some situations when only blood is assessed and thus the panel contends that desirable effects outweigh the undesirable effects. Most institutions have the capacity of obtaining these materials, which aids in feasibility, however, there will be some additional costs in having all suspected patients render a bone marrow aspirate. The cost will be minimal in the overall cost of care in patients will AL. The recommendation allows for the use of rendering a diagnosis on the blood only under certain circumstances, and thus the panel believes that the recommendation is feasible.

Statement 5. In addition to morphologic assessment (blood and bone marrow), the pathologist or treating clinician should obtain sufficient samples and perform conventional cytogenetic analysis (i.e., karyotype), appropriate molecular genetic and/or fluorescent in-situ hybridization (FISH) testing, and flow cytometric immunophenotyping (FCI). The flow cytometry panel should be sufficient to distinguish acute myeloid leukemia (including acute promyelocytic leukemia), T-cell acute lymphoblastic leukemia (T-ALL) (including early T-cell precursor leukemias), B-cell precursor ALL (B-ALL), and acute leukemia of ambiguous lineage on all patients diagnosed with acute leukemia. Molecular genetic and/or FISH testing does not, however, replace conventional cytogenetic analysis.

Note: If sufficient bone marrow aspirate or peripheral blood material is not available for FCI, immunohistochemical studies may be used as an alternative method for performing limited immunophenotyping. In addition, a second bone marrow core biopsy can be obtained and submitted unfixed in tissue culture media for disaggregation for genetic studies and flow cytometry.

Including cytogenetic analysis, appropriate molecular genetic and/or FISH testing and flow cytometric immunophenotyping will result in a more standardized initial workup which will likely reduce the need for repeat procedures/studies. Furthermore, these studies allow identification of parameters that may comprise a “fingerprint” of the leukemia and allow for detection of minimal residual disease in future specimens (see also Statement 12). Following the recommendation will increase costs to some degree. Most institutions have the capacity to perform this analysis or can send out to a reference laboratory when needed. The panel recognizes, however, that genetic testing might require additional resources.

Statement 6. For patients with suspected or confirmed acute leukemia, the pathologist may request and evaluate cytochemical studies to assist in the diagnosis and classification of acute myeloid leukemia (AML).

The panel believes that cytochemical studies play a minor role in the workup of AL, however, they are relatively simple to use and are inexpensive. The panel believes that while there is some utility, particularly for myeloperoxidase, cytochemical studies are becoming less useful in the diagnosis and classification of AL and that the evidence did not warrant this guideline statement to be raised to the level of a recommendation. Thus, this statement was designated an expert opinion.

Statement 7. The treating clinician or pathologist may use cryopreserved cells or nucleic acid, formalin fixed, non-decalcified paraffin-embedded (FFPE) tissue, or unstained marrow aspirate or peripheral blood smears obtained and prepared from peripheral blood, bone marrow aspirate or other involved tissues for molecular or genetic studies in which the use of such material has been validated. Such specimens must be properly identified and stored under appropriate conditions in a laboratory that is in compliance with regulatory and/or accreditation requirements.

Saving unused cells from the initial marrow procedure could circumvent an additional marrow procedure to obtain cells that might be necessary for additional diagnostic or prognostic studies, or to identify targets for therapy directed at specific antigens or genetic abnormalities. While fresh material is better, proper handling and storage of cells from the initial bone marrow (or peripheral blood cells in those cases where there are numerous neoplastic cells in the blood) should result in little to no undesirable effects. Improper handling/storage may result in false test results. Collecting marrow cells in formalin in a "clot section" takes only slightly more time at the bed side, whereas cryopreservation takes more time, special equipment, and facilities. This statement may therefore not be feasible for smaller institutions. Institutions that are a part of an AL cooperative group already use cryopreserved cells.

Statement 8. For patients with acute lymphoblastic leukemia (ALL) receiving intrathecal therapy, the treating clinician should obtain a cerebrospinal fluid (CSF) sample. The treating clinician or pathologist should ensure that a cell count is performed and that examination/enumeration of blasts on a cytocentrifuge preparation is performed and is reviewed by the pathologist.

Implementing this recommendation ensures that the pathologist has the specimens needed and the data from the specimen to inform an accurate diagnosis. The most common risks involved in obtaining a CSF are discomfort/pain, infection, and bleeding. However, the panel believes that obtaining a CSF as indicated in the guideline statement provides greater benefit than risk or harm. A CSF is relatively inexpensive and many pathologists already receive CSF samples in this context. The public comment period revealed that this guideline statement is acceptable.

Statement 9. For patients with acute leukemia other than those with ALL receiving intrathecal therapy, the treating clinician may, under certain circumstances, obtain a cerebrospinal fluid (CSF) sample when there is no clinical contraindication. The treating clinician or pathologist should ensure that a cell count is performed and that examination/enumeration of blasts on a cytocentrifuge preparation is performed and is reviewed by the pathologist.

Implementing this recommendation ensures that the pathologist has the specimens needed and the data from the specimen to inform an accurate diagnosis. Obtaining the sample at the time of intrathecal therapy means that the patient does not have to undergo a separate lumbar puncture procedure. The most common risks involved in obtaining a CSF are discomfort/pain, infection, and bleeding. However, the panel believes that obtaining a CSF as indicated in the guideline statement provides greater benefit than risk or harm, especially when performed at the time of intrathecal therapy. A CSF is relatively inexpensive and many pathologists already receive CSF samples in this context. The public comment period revealed that this guideline statement is acceptable.

Statement 10. For patients with suspected or confirmed acute leukemia, the pathologist may use flow cytometry in the evaluation of CSF.

The panel believes that this recommendation will help resolve morphologically difficult cases. While there is cost associated with flow cytometry, it is relatively small compared to overall care and many laboratories already perform this testing.

Statement 11. For patients who present with extramedullary disease without bone marrow or blood involvement, the pathologist should evaluate a tissue biopsy and process it for morphologic, immunophenotypic, cytogenetic, and molecular genetic studies, as recommended for the bone marrow.

Note: Additional biopsies may be indicated to obtain fresh material for ancillary testing.

Implementing this recommendation ensures that the pathologist has the specimens needed and the data from the specimens to inform an accurate diagnosis. Because obtaining a tissue biopsy is standard of care as are the laboratory methods indicated, implementing the recommendation is feasible in most institutions.

Statement 12. For patients with suspected or confirmed acute leukemia, the pathologist or treating clinician should ensure that flow cytometry analysis or molecular characterization is comprehensive enough to allow subsequent detection of minimal residual disease (MRD).

While this recommendation might require many institutions to change their procedures and/or workflow at increased expense, the panel believes that the desirable effects of being able to detect MRD outweigh the undesirable effects. The resources required to implement this recommendation will vary depending on the current set-up of the laboratory. The panel believes that this recommendation might encourage institutions to provide the resources to make this standard laboratory care.

Statement 13. For pediatric patients with suspected or confirmed B-ALL, the pathologist or treating clinician should ensure that testing for t(12;21)(p13.2;q22.1); *ETV6-RUNX1*, t(9;22)(q34.1;q11.2); *BCR-ABL1*, *KMT2A(MLL)* translocation, iAMP 21, and trisomy 4 and 10 is performed.

Statement 14. For adult patients with suspected or confirmed B-ALL, the pathologist or treating clinician should ensure that testing for t(9;22)(q34.1;q11.2) ; *BCR-ABL1* is performed. In addition, testing for *KMT2A* (*MLL*) translocations may be performed.

Statement 15. For patients with suspected or confirmed ALL, the pathologist or treating clinician may order appropriate mutational analysis for selected genes that influence diagnosis, prognosis, and/or therapeutic management that includes, but is not limited to: *PAX5, JAK1, JAK2,* and/or *IKZF1* for B-ALL and *NOTCH1* and/or *FBXW7* for T-ALL. Testing for overexpression of *CRLF2* may also be performed for B-ALL.

Statement 16. For pediatric and adult patients with suspected or confirmed acute myeloid leukemia (AML) of any type, the pathologist or treating clinician should ensure that testing for *FLT3*-ITD is performed. The pathologist or treating clinician may order mutational analysis that includes but is not limited to: *IDH1, IDH2, TET2, WT1, DNMT3A*, and/or *TP53* for prognostic and/or therapeutic purposes.

Statement 17. For adult patients with confirmed core binding factor (CBF) AML (AML with t(8;21)(q22;q22.1); *RUNX1-RUNX1T1* or inv(16)(p13.1q22) /t(16;16)(p13.1;q22); *CBFB-MYH11*, the pathologist or treating clinician should ensure that appropriate mutational analysis for *KIT* is performed. For pediatric patients with confirmed core binding factor AML (AML with t(8;21)(q22;q22.1); *RUNX1-RUNX1T1* or inv(16)(p13.1q22)

/t(16;16)(p13.1;q22); *CBFB-MYH11*, the pathologist or treating clinician may ensure that appropriate mutational analysis for *KIT* is performed.

These tests in Statements 13-17 all aid in improved prognosis determination and/or provide the treating clinician with information needed to make treatment decisions such as specific targeted therapy. The panel believes that the recommendations are acceptable as most of the respondents during public comment agreed with the statements. The panel does not suggest that this level of testing be performed at the average community laboratory where expertise or resources might be lacking, but instead believes that the resources required for laboratories already offering such complex testing will likely be small. One of the biggest benefits is that by providing these recommendations, pathologists who workup acute leukemia will be aware of the specific tests required for certain subsets of patients. Following the recommended tests in these guideline statements will require additional time to manage test ordering. The tests can also be costly and requires the laboratory personnel performing the tests to have molecular expertise.

Statement 18. For patients with suspected acute promyelocytic leukemia (APL), the pathologist or treating physician should also ensure that rapid detection of *PML-RARA* is performed. The treating physician should also order appropriate coagulation studies to evaluate for disseminated intravascular coagulation (DIC).

These tests all aid in improved prognosis determination and/or provide the treating clinician with information needed to make treatment decisions such as specific targeted therapy. The panel believes that the recommendations are acceptable as most of the respondents during public comment agreed with the statements. The panel does not suggest that this level of testing be performed at the average community laboratory where expertise or resources might be lacking, but instead believes that the resources required for laboratories already offering such complex testing will likely be small. One of the biggest benefits is that by providing these recommendations, pathologists who workup acute leukemia will be aware of the specific tests required for certain subsets of patients. Following the recommended tests in these guideline statements will require additional time to manage test ordering. The tests can also be costly and requires the laboratory personnel performing the tests to have molecular expertise.

Implementing this guideline statement will lead to the rapid diagnosis of APL, which is essential for the best outcome. The resources involved include altering laboratory workflow to accommodate a rush diagnosis, possible weekend staffing issues, and possible redundant tests being ordered to ensure a rapid result. The recommendation requires strong communication between the treating clinician, pathologist, and cytogenetics/molecular lab, which may not be in place in all settings.

Statement 19. For patients other than those with confirmed core binding factor AML, APL, or AML with myelodysplasia-related cytogenetic abnormalities, the pathologist or treating clinician should also ensure that mutational analysis for *NPM1, CEBPA*, and *RUNX1* is also performed.

These tests all aid in improved prognosis determination and/or provide the treating clinician with information needed to make treatment decisions such as specific targeted therapy. The panel believes that the recommendations are acceptable as most of the respondents during public comment agreed with the statements. The panel does not suggest that this level of testing be performed at the average community laboratory where expertise or resources might be lacking, but instead believes that the resources required for laboratories already offering such complex testing will likely be small. One of the biggest benefits is that by providing these recommendations, pathologists who workup acute leukemia will be aware of the specific tests required for certain subsets of patients. Following the recommended tests in these guideline statements will require additional time to manage test ordering. The tests can also be costly and requires the laboratory personnel performing the tests to have molecular expertise.

Implementing this guideline statement will provide data used to make prognostic/therapeutic decisions. The panel believes that the value of these tests outweighs the risks or harms, which are negligible assuming that the initial specimen collection yields enough material to perform the additional testing. Implementing this guideline statement may require extra time from the pathologist or clinician to manage testing orders.

Statement 20. For patients with confirmed acute leukemia, no recommendation is made for or against the use of global/gene specific methylation, micro RNA (miRNA) expression, or gene expression analysis for diagnosis or prognosis.

These tests all aid in improved prognosis determination and/or provide the treating clinician with information needed to make treatment decisions such as specific targeted therapy. The panel believes that the recommendations are acceptable as most of the respondents during public comment agreed with the statements. The panel does not suggest that this level of testing be performed at the average community laboratory where expertise or resources might be lacking, but instead believes that the resources required for laboratories already offering such complex testing will likely be small. One of the biggest benefits is that by providing these recommendations, pathologists who workup acute leukemia will be aware of the specific tests required for certain subsets of patients. Following the recommended tests in these guideline statements will require additional time to manage test ordering. The tests can also be costly and requires the laboratory personnel performing the tests to have molecular expertise.

Although there is evidence that these tests influence AML prognosis, currently they are not actionable with respect to treatment strategy and are not readily available in most accredited diagnostic laboratories. The resources to implement and pay for this testing are large and would require a large investment. Based on the evidence and expert experience, the panel is uncertain as to whether there is great benefit versus harm and thus could not make a recommendation for or against the use of these tests.

Statement 21. For patients with confirmed mixed phenotype acute leukemia (MPAL), the pathologist or treating clinician should ensure that testing for t(9;22)(q34.1;q11.2); *BCR-ABL1*, and *KMT2A* (*MLL*) translocations is performed.

These tests all aid in improved prognosis determination and/or provide the treating clinician with information needed to make treatment decisions such as specific targeted therapy. The panel believes that the recommendations are acceptable as most of the respondents during public comment agreed with the statements. The panel does not suggest that this level of testing be performed at the average community laboratory where expertise or resources might be lacking, but instead believes that the resources required for laboratories already offering such complex testing will likely be small. One of the biggest benefits is that by providing these recommendations, pathologists who workup acute leukemia will be aware of the specific tests required for certain subsets of patients. Following the recommended tests in these guideline statements will require additional time to manage test ordering. The tests can also be costly and requires the laboratory personnel performing the tests to have molecular expertise.

Statement 22. All laboratory testing performed for the initial workup and diagnosis of a patient with acute leukemia must be performed in a laboratory that is in compliance with regulatory and/or accreditation requirements.

The statement codifies current laboratory practice in the United States (US), and speaks against diagnostic testing being performed in research laboratories. Thus, this statement is necessary (as mandated by the Clinical Laboratory Improvement Amendments [CLIA] and beneficial for ensuring high quality patient care, but probably not of high impact with respect to current practice, particularly in the US.

Statement 23. If after examination of a peripheral blood smear, it is determined that the patient will require immediate referral to another institution with expertise in the management of acute leukemia for treatment, the initial institution should, whenever possible, defer invasive procedures including bone marrow aspiration and biopsies to the treatment center to avoid duplicate procedures, associated patient discomfort, and additional costs.

Deferring invasive procedures should result in better, coordinated care, increased comfort for the patient, and in cost savings. The recommendation benefits patients in that multiple bone marrows and/or biopsies would be avoided. However, if the clinical context of this recommendation is misinterpreted, it could result in delayed care. The panel does not suggest that care be delayed or prohibited; however, the panel strongly believes that patients will benefit from having the treating center take the primary role in performing invasive procedures. This recommendation will require institutions to determine the level of care it will provide patients being worked up for acute leukemia.

Statement 24. If a patient is referred to another institution for treatment, the primary institution should provide the treatment center with all laboratory results, pathology slides, flow cytometry data, cytogenetic information, and a list of pending tests at the time of the referral. Pending test results should be forwarded as they become available.

The benefits of this recommendation are that treatment centers will be have the information and laboratory assets necessary to properly treat patients including knowledge of tests ordered, but have information regarding which results are still pending. This should result in improved coordinated care and cost savings. The primary institution is charged with collating and providing these items to the treatment center, although in the vast majority of cases the recommendation codifies current practices.
There is the possibility that slides in transit could be lost, or waiting for slides could delay patient care. Resources required would include cost of copying/mailing slides and monitoring pending results. Shared medical records may enhance this capability with minimal cost.

Statement 25. In the initial report, the pathologist should include laboratory, morphologic, immunophenotypic, and, if performed, cytochemical data, on which the diagnosis is based, along with a list of any pending tests. The pathologist should issue addenda/amended reports when the results of additional tests become available.

The benefits of implementing this guideline statement are that the laboratory report will contain the results for each of the elements used to render a diagnosis and that this information will be visible to both laboratory personnel and treating clinicians. This includes a list of pending tests which will inform both parties that additional laboratory data is forthcoming. The panel could not identify any undesirable effects associated with this guideline statement. The feedback received during the public comment period showed that almost all respondents (161 of 171) agreed with this guideline statement.

Statement 26. The pathologist and treating clinician should coordinate and ensure that all tests performed for classification, management, predicting prognosis and disease monitoring are entered into the patient's medical records.

Note: This information should include the sample source, adequacy, and collection information as applicable.

Implementing this guideline statement should improve patient care by ensuring that the treating physicians have all available information with an appropriate and integrated interpretation. The panel recognizes that the time required to enter the information into the patient's medical records puts additional burden on the pathologist. The panel also acknowledges that this will require electronic reporting solutions or additional man power which might require substantial resources.

Currently most laboratories or hospital information systems do not easily allow for entering these elements into the patient's medical records, but the panel contends that doing so would almost certainly improve patient care. With the advent of electronic medical records, entering this information should become easier in the foreseeable future.

Statement 27. Treating physicians and pathologists should use the current World Health Organization (WHO) terminology for the final diagnosis and classification of acute leukemia.

Since many institutions already look to the WHO for international classification standards, there is little, if no harm, in using this terminology for the final diagnosis and classification of AL. The major benefit of using the WHO schema is that physicians will have a consistent understanding of diagnosis and classification of AL regardless of geographic parameters.

Supplemental Table 1: Summary of Data for Patient History

Author, year	Year	Study design	Influence of history of Down Syndrome	Influence of family history of leukemia
Gale RE et al[50]	2008	PCS	-	In MVA for CR, de novo/secondary OR=2.39 (95% CI:1.33-4.29) *P*=.01; for OS OR=1.37 (95% CI:1.04-1.80) *P*=.02 N=116
Wandt H et al[48]	2008	PCS	-	History of AML (de novo versus secondary) was significantly associated with CR: OR=0.57 (95% CI:.42-.76), *P*<.001 N=379
Lugthart S et al[11]	2010	PCS	-	Secondary AML versus de novo AML: OR of achieving CR=0.56, *P*=.01; HR of OS:1.27, *P*=.001; HR of EFS=1.23, *P*=.01 N=272
Maloney KW et al[15]	2010	PCS	5 year OS for DS=85.8% ± 6.5% versus 90.% ± 0.9% for non-DS ALL, *P*=.03 N=80	

ALL – acute lymphoblastic leukemia; AML – acute myeloid leukemia; CI – confidence interval; CR – complete response; DS – Down Syndrome; EFS – event-free survival; HR – hazard ratio; MVA – multivariate analysis; PCS – prospective cohort study; OR – odds ratio; OS – overall survival;

Supplemental Table 2: Summary of Data for Ethnicity and Performance Status

Author, year	Year	Study design	Relevance of ethnicity to diagnosis of AL	Relevance of ethnicity to outcome in AL	Relevance of performance status to outcome of AL
Gale RE et al[50]	2008	PCS			WHO performance status (in MVA CR: OR=1.43 (95% CI:1.19-1.71) *P*<.001; OS: OR=1.13 (95% CI:1.05-1.22) *P*=.01
Maloney KW et al[15]	2010	PCS	Distribution of race was significantly different between the 2 groups (*P*<.001), with children with DS-ALL being largely		

Author, year	Year	Study design	Relevance of ethnicity to diagnosis of AL	Relevance of ethnicity to outcome in AL	Relevance of performance status to outcome of AL
			white (91.2%) with no black and 8.8% "other." In contrast, children with non–DS-ALL were 76.4% white, 7.7% black, and 15.9% other.		
Salzer WL et al[26]	2010	PCS		Yes (study provided no details or data)	
Harvey RC et al[85]	2010	PCS		Hispanic ethnicity is associated with worse 4-year relapse free survival (*P*<.001)	

AL – acute leukemia; ALL – acute lymphoblastic leukemia; CI – confidence interval; CR – complete response; DS – Down Syndrome; MVA – multivariate analysis; OR – odds ratio; PCS – prospective cohort study; WHO – World Health Organization;

Supplemental Table 3: Summary of Data for Relevance of CNS Symptoms for Patients with ALL

Author, year	Year	Study design	Relevance of CNS symptoms on outcomes for patients with ALL
Salzer WL et al[26]	2010	PCS	Positive CNS status in non-infant B-ALL was associated with poor outcomes HR=1.34 *P*=.04
Schmiegelow K et al[27]	2010	PCS	ALL-92 study: Patients with CNS3 had poorer outcomes than patients with CNS1, CNS2 EFS: *P*=.01; OS: *P*=.001; ALL-2000 study: EFS: *P*=.01, OS: *P*=.02
Pui CH et al[28]	2010	PCS	Patients with CNS3 had poorer outcomes than those with CNS1, CNS2 CNS status in Study 11:

Author, year	Year	Study design	Relevance of CNS symptoms on outcomes for patients with ALL
			EFS: *P*<.001 OS: *P* =.01 Study 12: EFS: *P* =.03 OS: *P* =.22
Scrideli CA et al[36]	2009	PCS	Patients with CNS3 had poorer outcomes than those with CNS1, CNS2 5-year EFS: CNS3: 79.4 ± 3.1% CNS1 or CNS2: 40 ± 17.4%; *P* <.001

ALL – acute lymphoblastic leukemia; B-ALL – B- cell precursor acute lymphoblastic leukemia; CNS – central nervous system; EFS – event-free survival; HR – hazard ratio; OS – overall survival; PCS – prospective cohort study;

Supplemental Table 4: Summary of Data for Gene Specific Methylation and miRNA Expression

Author, year	Year	Study design	Gene specific methylation	miRNA expression	Other
Damm F et al[116]	2012	NRCT			*ID1*, *P*=NS for OS in MVA. *BAALC* high expression, *P*=.001 for inferior OS in MVA in CN-AML
Schwind S et al[119]	2011	NRCT			Low *MN1* expressors had longer OS, and EFS than high *MN1* expressors (OS 1.2 years versus .8 years *P*=.03 and EFS .7 years versus .2 years *P*=.01) in CN-AML
Alvarez S et al[68]	2010	PCS	*BRINP1/DBC1*, *P*=NS for OS and EFS in MVA in CN-AML		
Langer C et al[45]	2008	PCS			Higher *BAALC* expression associated with shorter OS (*P*=.04), but not DFS (*P*=NS) in MVA of CN-AML. *ERG P*=NS for DFS and OS in MVA.
Schwind S	2010	PCS		Higher miR-181a associated	

Author, year	Year	Study design	Gene specific methylation	miRNA expression	Other
et al[70]				with longer OS and EFS in MVA of CN-AML, *P*<.05	
Marucci G et al,[75] 2008	2008	PCS		Higher miR-181a, 181b associated with more favorable outcome, Higher miR124, 128, 194, 219, 220a, 320 associated with worse outcome in CN-AML, all *P*<.05	
Schwind S et al[103]	2010	PCS			Low *BAALC* and *ERG* expression in older CN-AML patients resulted in longer DFS (*P*=.01 and *P*=.03) and OS (*P*<.001 and *P*=.01) respectively.
Chuang M-K et al[120]	2015	PCS		hsa-miR-9-5p and hsa-miR-155-5p associated with worse OS (P=.01 and *P*=.03 respectively). hsa-miR-203 trend of association with favorable OS (*P*=.08). Score using these 3 miRNA was independently associated with OS in test and validation de novo AML patient cohorts (*P*<.001 for both).	
Roman-Gomez J et al[101]	2009	PCS		Examined methylation profile of genes of 9 miRNA in ALL, methylation of any one or more was associated with shorter OS and EFS on MVA.	

BAALC – Brain And Acute Leukemia, Cytoplasmic; CN-AML – cytogenetically normal acute myeloid leukemia; *BRINP1/DBC1* – Bone Morphogenetic Protein/Retinoic Acid Inducible Neural-Specific 1/Deleted in Bladder Cancer Protein 1; DFS – disease-free survival; EFS – event-free survival; *ERG* – ETS transcription factor gene; HR – hazard ratio; hsa-miR – human messenger ribonucleic acid; *ID1* – Inhibitor Of DNA Binding 1; mRNA – messenger ribonucleic acid; miRNA – micro ribonucleic acid; *MN1* – Meningioma 1; MVA – multivariate analysis; NRCT – non-randomized control trial; NS – not significant; OS – overall survival; PCS – prospective cohort study

Supplemental Table 5: Summary of Data for Location of Laboratory Testing, Differences in Results, Description of Specimen Transport and Cryopreservation Technique

Author, year	Year	Study Design	Were all laboratory studies performed at the primary institution ?	Were some laboratory studies performed at a reference laboratory, central protocol laboratory, or tertiary care center?	If duplicate laboratory studies were performed at two different institutions were there significant differences in the results?	If laboratory studies were referred to a central, reference, or tertiary care laboratory how were the specimens transported ?	If laboratory studies were referred to a central, reference, or tertiary care laboratory were the samples cryopreserved ?	Description of cryopreservation technique
Krauter J et al[37]	2009	Meta-analyis (M/A)	No (N)	Yes (Y)	NR (Not reported)	NR	NR	NR
Schaich M et al[51]	2007	M/A	N	Y	NR	NR	NR	NR
Vance G et al[109]	2007	Randomized controlled trial (RCT)	N	Y	NR	NR	NR	NR
Moorman AV et al[18]	2010	Non-randomized Control Trial (N-RCT)	N	Y	NR	NR	NR	NR
Arico M et al[47]	2008	N-RCT	N	Y	NR	NR	NR	NR
Metzeler K et al[4]	2011	Prospective Cohort Study (PCS)	N	Y	NR	NR	NR	NR
Montesinos P et al[6]	2011	PCS	N	Y	N	NR	NR	NR
Kayser S et al[8]	2011	PCS	N	Y	N	NR	NR	NR

Clappier E et al[9]	2010	PCS	N	Y	N	NR	NR	NR
Rollig C et al[14]	2010	PCS	Y	N	N	NR	NR	NR
Grimwade D et al[19]	2010	PCS	N	Reference laboratory only	NR	NR	NR	NR
Basso G et al[31]	2009	PCS	N	Reference laboratory only	N	Bone marrow samples shipped over night to reference laboratory	NR	NR
Scrideli CA et al[36]	2009	PCS	Y	N	N	NR	NR	NR
Lo-Coco F et al[42]	2008	PCS	N	Y	N	NR	NR	NR
Forestier E et al[127]	2008	PCS	N	Y	N	NR	NR	NR
Wandt H et al[48]	2008	PCS	Y	N	Not applicable (NA)	NR	NR	NR
Barbaric D et al[54]	2007	PCS	N	Y	Y	NR	NR	NR
Mrózek K et al[57]	2008	PCS	N	Y	Y	NR	NR	NR
Pabst T et al[60]	2009	PCS	Y	N	NR	NR	NR	NR
Marucci G et al[61]	2008	PCS	N	Y	N	NR	NR	NR
Abbas S et al[71]	2010	PCS	Y	N	NR	NR	NR	NR
Whitman SP et al[76]	2008	PCS	N	Y	N	NR	Y	Trizol Reagent; Invitrogen, Carlsbad, CA
Chiaretti S et al[80]	2007	PCS	N	Reference laboratory only	NR	NR	Y	Liquid nitrogen
Fischer L et al[83]	2009	PCS	N	Y	N	NR	NR	NR
Medeiros BC et al[94]	2010	PCS	Y	N	NR	NR	NR	NR

Milani L et al[96]	2010	PCS	N	Reference laboratory only	N	NR	Y	2-10 million cells immediately frozen and stored at -70 in tissue banks
Moorman AV et al[108]	2007	PCS	Varied	Varied	N	NR	NR	NR
Patel JL et al[111]	2012	PCS	N	Y	N	NR	NR	NR

Supplemental Table 6: Quality Assessment Criteria by Study Design

Criteria	**Study Design**			
	Clinical Practice Guideline/Systematic Review (CPG/SR)	**Meta-analysis**	**Randomized Control Trial (RCT)/Quasi Randomized Control Trial (QRCT)**	**Non-randomized Control Trial (NRCT)/Prospective Cohort Study (PCS) /RCS**
Based on a systematic review	✓ (CPG only)	✓		
Included a multidisciplinary panel	✓			
Patient preferences were considered	✓			
Important patient sub-types were considered	✓			
Methods were well-described and reproducible	✓	✓		
Information on potential conflicts of interest were gathered and disclosed	✓			
Quality of the evidence was assessed	✓	✓		

Strength of the evidence was rated	✓			
CPG includes a plan for updating	✓			
Sources of funding are disclosed	✓	✓	✓	✓
Any planned pooling was stated a priori		✓		
Limitations of the analysis are discussed		✓		
Randomization method fully-described			✓	
Details on any blinding was provided			✓	
Provided details of all planned analyses			✓	
Stated the expected effect size and described the statistical power calculation			✓	
Reported the length of follow-up			✓	
Provided a description of the baseline characteristics for all patients by treatment/assessment arm			✓	✓
Balance between treatment/assessment groups				✓
Reporting if any adjustments were made where baseline differences were detected				✓

√ indicates that for a specified study design (per the column header) an assessment of the listed criteria (per row header) was included in the quality assessment

Supplemental Table 7: Quality Assessment for Statement 1

Non-randomized Control Trials (N=2)						
Author	**Year**	**Was there balance between treatment/assessment groups?**	**Reporting of baseline characteristics (and any differences detected between groups)**	**Reporting of any adjustment when differences were present**	**Funding source**	**Overall risk of bias assessment**
Mendler JH et al[115]	2012	Single cohort	Not reported (NR)	NR	Non-industry	Low-moderate
Damm F et al[116]	2012	Single cohort	NR	NR	Non-industry	Low-moderate
Prospective Cohort Studies (N=28)						
Author	**Year**	**Was there balance between treatment/assessment groups?**	**Reporting of baseline characteristics (and any differences detected between groups)**	**Reporting of any adjustment when differences were present**	**Funding source**	**Overall risk of bias assessment**
Kuhnl A et al[23]	2010	Single cohort	Yes (Y)	Not applicable (N/A)	Non-industry	Low-moderate
Pui CH et al[28]	2010	No (N) (325:188:165:247:53)	Y	No (N)	Non-industry	Low-moderate
Escherich G et al[25]	2010	N (129:289:205:519:667)	Y	NR	Non-industry	Low-moderate
Salzer WL et al[26]	2010	Single cohort	Y	N/A	Non-industry	Low-moderate

Marks DI et al[30]	2009	Single cohort	Y	N/A	Non-industry	Low-moderate
Scrideli CA et al[36]	2009	Single cohort	Y	N/A	NR	Moderate
Roman-Gomez J et al[55]	2007	Single cohort	Y	N/A	Non-industry	Low-moderate
Tauchi H et al[105]	2008	Single cohort	Y	N/A	Non-industry	Low-moderate
Roman-Gomez J et al[101]	2009	Single cohort	Y	N/A	Non-industry	Low-moderate
Moorman AV et al[114]	2012	Single cohort	N	N/A	Non-industry	Low-moderate
Taskesen E et al[5]	2011	N (1031:60:91)	Y	Y	Non-industry	Low
Lugthart S et al[11]	2010	N (94:52:19:123:2231)	Y	NR	Non-industry	Low-moderate
Rollig C et al[14]	2010	Single cohort	Y	N/A	NR	Moderate

Wagner K et al[21]	2010	Single cohort	Y	N/A	Non-industry	Low-moderate
Groschel S et al[22]	2010	N (458:924)	Y	N	Non-industry	Low-moderate
Lo-Coco F et al[42]	2008	Single cohort	N	N/A	Non-industry	Low-moderate
Langer C et al[45]	2008	Single cohort	Y	N/A	Non-industry	Low
Wandt H et al[48]	2008	Single cohort	Y	N/A	Non-industry	Low-moderate
Gale RE et al[50]	2008	Single cohort	Y	N/A	Non-industry	Moderate
Dufour A et al[58]	2010	Single cohort	Y	N/A	Non-industry	Low-moderate
Damm F et al[59]	2010	N (249:50)	Y	N	NR	Moderate
Schwind S et al[70]	2010	Y (187:122)	Y	Y	Non-industry	Low
Santamaria C et al[73]	2010	Single cohort	Y	N/A	Non-industry	Low-moderate

Seifert H et al[74]	2009	Single cohort	Y	N/A	NR	Moderate
Medeiros BC et al[94]	2010	Single cohort	Y	N/A	NR	Moderate
Schneider F et al[118]	2012	Single cohort	Y	N/A	Partial industry	Low-moderate
Maloney KW et al[15]	2010	N (80:2731)	Y	NR	Non-industry	Low-moderate
Harvey RC et al[85]	2010	Single cohort	Y	N/A	Partial industry	Moderate

Supplemental Table 8: Quality Assessment for Statement 2

Author	**Year**	**Was there balance between treatment/assessment groups?**	**Reporting of baseline characteristics (and any differences detected between groups)**	**Reporting of any adjustment when differences were present**	**Funding source**	**Overall risk of bias assessment**
Salzer WL et al[26]	2010	Single cohort	Yes (Y)	Not applicable (N/A)	Non-industry	Low-moderate
Schmiegelow K et al[27]	2010	No (N) (1645:1023)	Y	Y	Non-industry	Low
Pui CH et al[28]	2010	N (325:188:165:247:53)	Y	No	Non-industry	Low-moderate

Scrideli CA et al[36]	2009	Single cohort	Y	N/A	Not reported	Moderate

Supplemental Table 9: Quality Assessment for Statement 3

Author	Year	Provided details on randomization	Provided details on blinding	Provided details on any planned analysis	Expected effect size calculation and power calculation	Reported on length of follow-up	Reported on any differences in patient characteristics	Funding source	Overall risk of bias assessment
Randomized Controlled Trials (N=2)									
Lange BJ et al[49]	2008	Not reported (NR), but was randomized	NR	Yes (Y)	Y	Y, 56 month median	Y	Non-industry	Very low
Schneider F et al[62]	2009	NR, but was randomized	NR	NR	NR	NR	Y	Non-industry	High

Non-randomized Controlled Trials (N=5)						
Author	**Year**	**Was there balance between treatment/assessment groups?**	**Reporting of baseline characteristics (and any differences detected between groups)**	**Reporting of any adjustment when differences were present**	**Funding source**	**Overall risk of bias assessment**
Moorman AV et al[18]	2010	NR	NR	NR	Non-industry	Moderate
Oudot C et al[46]	2008	No (N) (53:1333)	Y	NR	Non-industry	Low-moderate
Arico M et al[47]	2008	N (115:1385:244:1744)	Y	NR	Non-industry	Low-moderate
Damm F et al[116]	2012	Single cohort	NR	NR	Non-industry	Low-moderate

Schwind S et al[119]	2011	Single cohort	NR	NR	Non-industry	Low-moderate
Prospective Cohort Studies (N=44)						
Gaidzik VI et al[3]	2011	Single cohort	Y	Not applicable (N/A)	Non-industry	Low-moderate
Metzeler K et al[4]	2011	N (104:323)	Y	Y	Non-industry	Low
Taskesen E et al[5]	2011	N (1031:60:91)	Y	Y	Non-industry	Low
Montesinos P et al[6]	2011	N (72:579)	Y	Y	Non-industry	Low
Stolzel F et al[7]	2011	N (233:72)	Y	NR	Non-industry	Low-moderate
Kayser S et al[8]	2011	N (200:2653)	Y	NR	Non-industry	Low-moderate
Tallman MS et al[10]	2010	N (174:175:554)	Y	No (N)	Non-industry	Low-moderate
Lugthart S et al[11]	2010	N (94:52:19:123:2231)	Y	NR	Non-industry	Low-moderate
Rollig C et al[14]	2010	Single cohort	Y	N/A	NR	Moderate
Ho PA et al[17]	2010	Single cohort	Y	N/A	Non-industry	Low-moderate
Wagner K et al[21]	2010	Single cohort	Y	N/A	Non-industry	Low-moderate
Groschel S et al[22]	2010	N (458:924)	Y	N	Non-industry	Low-moderate

Kuhnl A et al[23]	2010	Single cohort	Y	N/A	Non-industry	Low-moderate
Escherich G et al[25]	2010	N (129:289:205:519:667)	Y	NR	Non-industry	Low-moderate
Salzer WL et al[26]	2010	Single cohort	Y	N/A	Non-industry	Low-moderate
Schmiegelow K et al[27]	2010	N (1645:1023)	Y	Y	Non-industry	Low
Pui CH et al[28]	2010	N (325:188:165:247:53)	Y	N	Non-industry	Low-moderate
Marks DI et al[30]	2009	Single cohort	Y	N/A	Non-industry	Low-moderate
Metzeler KH et al[32]	2009	Single cohort	Y	N/A	Non-industry	Low-moderate
Karrman K et al[35]	2009	Single cohort	Y	N/A	Non-industry	Low-moderate
Renneville A et al[117]	2012	Single cohort	Y	N/A	Non-industry	Low-moderate
Gaidzik VI et al[38]	2009	Single cohort	Y	N/A	Non-industry	Low-moderate
Virappane et al[40]	2008	Single cohort	Y	N/A	Non-industry	Low-moderate
Lo-Coco F et al[42]	2008	Single cohort	N	N/A	Non-industry	Low-moderate
Langer C et al[45]	2008	Single cohort	Y	N/A	Non-industry	Low

Gale RE et al[50]	2008	Single cohort	Y	N/A	Non-industry	Moderate
Yanada M et al[53]	2007	Single cohort	Y	N/A	Non-industry	Low-moderate
Damm F et al[59]	2010	N (249:50)	Y	N	NR	Moderate
Pabst T et al[60]	2009	Single cohort	N	N/A	Non-industry	Low-moderate
Marcucci G et al[61]	2008	Single cohort	Y	N/A	Non-industry	Low-moderate
Paschka P et al[63]	2008	Single cohort	Y	N/A	Non-industry	Low-moderate
Damm F et al[64]	2011	Y (275:293)	Y	None required	Non-industry	Low
Becker H et al[65]	2010	Single cohort	Y	N/A	Non-industry	Low-moderate
Rubnitz JE et al[67]	2007	Single cohort	Y	N/A	Non-industry	Low-moderate
Schwind S et al[70]	2010	Y (187:122)	Y	Y	Non-industry	Low
Santamaria C et al[73]	2010	Single cohort	Y	N/A	Non-industry	Low-moderate
Jiao B et al[89]	2009	Single cohort	Y	N/A	Non-industry	Low-moderate
Johnston DL et al[90]	2010	N (1113:192:154:148)	NR	NR	Non-industry	Moderate
Moorman AV	2007	Single cohort	Y	N/A	Non-industry	Low-moderate

et al[97]						
Prebet T et al[100]	2009	Single cohort	Y	N/A	NR	Moderate
Santamaria CM et al[102]	2009	Single cohort	Y	N/A	Non-industry	Low-moderate
Schwind S et al[103]	2010	Single cohort	Y	N/A	Non-industry	Low-moderate
Zachariadis V et al[107]	2011	Single cohort	Y	N/A	Non-industry	Low-moderate
Wheatley K et al[110]	2009	Single cohort	Y	N/A	NR	Moderate

Supplemental Table 10: Quality Assessment for Statement 7

Prospective Cohort Studies (N=4)						
Author	**Year**	**Was there balance between treatment/assessment groups?**	**Reporting of baseline characteristics (and any differences detected between groups)**	**Reporting of any adjustment when differences were present**	**Funding source**	**Overall risk of bias assessment**
Marcucci G et al[75]	2008	Yes (Y) (65:55)	Y	None required	Non-industry	Low
Whitman SP et al[76]	2008	No (N) (123:16)	Y	Y	Non-industry	Low
Hollink IH et al[87]	2009	Single cohort	Y	Not applicable (N/A)	Non-industry	Low-moderate
Milani L	2010	Single cohort	Y	N/A	Non-industry	Low-moderate

et al[96]						

Supplemental Table 11: Quality Assessment for Statement 12

Author	Year	Provided details on randomization	Provided details on blinding	Provided details on any planned analysis	Expected effect size calculation and power calculation	Reported on length of follow-up	Reported on any differences in patient characteristics	Funding source	Overall risk of bias assessment
Yin JAL et al[112]	2012	Not reported (NR), but was randomized	NR	Yes (Y)	NR	Y, 36 month median	Y	Non-industry	Moderate

Author	Year	Was there balance between treatment/assessment groups?	Reporting of baseline characteristics (and any differences detected between groups)	Reporting of any adjustment when differences were present	Funding source	Overall risk of bias assessment
Zhou J et al[52]	2007	Single cohort	Y	Not applicable (N/A)	Non-industry	Low
Waanders E et al[106]	2011	Single cohort	No (N)	N/A	Non-industry	Low-moderate
Basso G et al[31]	2009	Single cohort	Y	N/A	Non-industry	Low-moderate
Maloney KW et al[15]	2010	N (80:2731)	Y	NR	Non-industry	Low-moderate
Escherich G et al[25]	2010	N (129:289:205:519:667)	Y	NR	Non-industry	Low-moderate
Schmiegelow K et	2010	N (1645:1023)	Y	Y	Non-industry	Low

al[27]						
Pui CH et al[28]	2010	N (325:188:165:247:53)	Y	N	Non-industry	Low-moderate
Scrideli CA et al[36]	2009	Single cohort	Y	N/A	NR	Moderate
Mullighan CG et al[98]	2009	Y (221:258)	Y	N/A	Partial industry	Low-moderate
Markova J et al[33]	2009	Single cohort	Y	N/A	Non-industry	Low-moderate
Buccisano F et al[12]	2010	Single cohort	Y	N/A	NR	Moderate
Maurillo L et al[93]	2008	Single cohort	Y	N/A	Non-industry	Low-moderate
Patel B et al[29]	2010	Single cohort	Y	N/A	Non-industry	Low-moderate
Jeha S et al[88]	2009	Single cohort	N	N/A	Non-industry	Low-moderate

Supplemental Table 12: Quality Assessment for Statement 13

Prospective Cohort Studies (N=6)						
Author	**Year**	**Was there balance between treatment/assessment groups?**	**Reporting of baseline characteristics (and any differences detected between groups)**	**Reporting of any adjustment when differences were present**	**Funding source**	**Overall risk of bias assessment**
Escherich G	2010	No (N)	Yes (Y)	Not reported	Non-	Low-

et al[25]		(129:289:205:519:667)		(NR)	industry	moderate
Salzer WL et al[26]	2010	Single cohort	Y	Not applicable (N/A)	Non-industry	Low-moderate
Schmiegelow K et al[27]	2010	N (1645:1023)	Y	Y	Non-industry	Low
Rubnitz JE et al[44]	2008	Single cohort	Y	N/A	Non-industry	Low
Pui CH et al[28]	2010	N (325:188:165:247:53)	Y	N	Non-industry	Low-moderate
Moorman AV et al[97]	2007	Single cohort	Y	N/A	Non-industry	Low-moderate

Supplemental Table 13: Quality Assessment for Statement 14

Prospective Cohort Studies (N=2)						
Author	**Year**	**Was there balance between treatment/assessment groups?**	**Reporting of baseline characteristics (and any differences detected between groups)**	**Reporting of any adjustment when differences were present**	**Funding source**	**Overall risk of bias assessment**
Moorman AV et al[108]	2007	Single cohort	Yes (Y)	Not applicable (N/A)	Non-industry	Low-moderate

Kuhnl A et al[23]	2010	Single cohort	Y	N/A	Non-industry	Low-moderate

Supplemental Table 14: Quality Assessment for Statement 15

Prospective Cohort Studies (N=11)						
Author	**Year**	**Was there balance between treatment/assessment groups?**	**Reporting of baseline characteristics (and any differences detected between groups)**	**Reporting of any adjustment when differences were present**	**Funding source**	**Overall risk of bias assessment**
Mullighan CG et al[98]	2009	Yes (Y) (221:258)	Y	Not applicable (N/A)	Partial industry	Low-moderate
Cario G et al[79]	2010	Single cohort	Y	N/A	Non-industry	Low-moderate
Flex E et al[84]	2008	No (N) (88:38:85:49)	N	Not reported (NR)	Non-industry	Moderate
Moorman AV et al[114]	2012	Single cohort	N	N/A	Non-industry	Low-moderate
Familiades J et al[34]	2009	Single cohort	Y	N/A	Non-industry	Low-moderate
Den Boer ML et al[122]	2009	N (190:107)	Y	NR	Non-industry	Low-moderate

Kuiper RP et al[69]	2010	N (102:29)	N	NR	Non-industry	Moderate
Harvey RC et al[85]	2010	Single cohort	Y	N/A	partial industry	Moderate
Clappier E et al[9]	2010	Single cohort	Y	N/A	Non-industry	Low-moderate
Marks DI et al[30]	2009	Single cohort	Y	N/A	Non-industry	Low-moderate
Asnafi V et al[56]	2009	Single cohort	Y	N/A	Non-industry	Low-moderate
Baldus CD et al[77]	2009	Single cohort	Y	N/A	Non-industry	Low-moderate
Heesch S et al[86]	2010	Single cohort	Y	N/A	Non-industry	Low-moderate
Kox C et al[92]	2010	N (301:151)	N	Y	Non-industry	Low

Supplemental Table 15: Quality Assessment for *FLT3*-ITD Studies

Prospective Cohort Studies (N=13)						
Author	**Year**	**Was there balance between treatment/assessment groups?**	**Reporting of baseline characteristics (and any**	**Reporting of any adjustment when differences**	**Funding source**	**Overall risk of bias assessment**

			differences detected between groups)	**were present**		
Gaidzik VI et al[3]	2011	Single cohort	Yes (Y)	Not applicable (N/A)	Non-industry	Low-moderate
Kayser S et al[8]	2011	No (N) (200:2653)	Y	Not reported (NR)	Non-industry	Low-moderate
Buccisano F et al[12]	2010	Single cohort	Y	N/A	NR	Moderate
Rollig C et al[14]	2010	Single cohort	Y	N/A	NR	Moderate
Groschel S et al[22]	2010	No (N) (458:924)	Y	NR	Non-industry	Low-moderate
Kayser S et al[91]	2009	Single cohort	Y	N/A	Non-industry	Low-moderate
Ho PA et al[17]	2010	Single cohort	Y	N/A	Non-industry	Low-moderate
Abbas S et al[71]	2010	Single cohort	Y	N/A	Non-industry	Low-moderate
Ho PA et al[20]	2010	Single cohort	Y	N/A	Non-industry	Low-moderate
Pollard JA et al[24]	2010	Single cohort	Y	N/A	Non-industry	Low-moderate
Markova J et al[33]	2009	Single cohort	Y	N/A	Non-industry	Low-moderate
Jiao B et al[89]	2009	Single cohort	Y	N/A	Non-industry	Low-moderate

Schlenk RF et al[121]	2014	Y (81:80:81:81)	Y	N/A	Non-industry	Low

Supplemental Table 16: Quality Assessment for All Other Mutational Testing in AML (Excluding *FLT3*-ITD)

Author	Year	Based on systematic review	Reproducible Methods	Quality assessment of included studies	Planned pooling stated a priori	Limitations of the study	Funding source	Overall risk of bias assessment
Meta-analysis (N=1)								
Zhou KG et al[125]	2012	Yes (Y)	Y	Y	Y	Y	Non-industry	Low

Author	Year	Was there balance between treatment/assessment groups?	Reporting of baseline characteristics (and any differences detected between groups)	Reporting of any adjustment when differences were present	Funding source	Overall risk of bias assessment
Non-randomized Controlled Trials (N=3)						
Nomdedeu J et al[126]	2012	Single cohort	Not reported (NR)	NR	Non-industry	Low-moderate
Mendler JH e al[115]	2012	Single cohort	NR	NR	Non-industry	Low-moderate
Damm F et al[116]	2012	Single cohort	NR	NR	Non-industry	Low-moderate
Prospective Cohort Studies (N=17)						
Paschka P et al[123]	2010	Single cohort	Yes (Y)	Not applicable (N/A)	Non-industry	Low-moderate
Abbas S et al[71]	2010	Single cohort	Y	N/A	Non-industry	Low-moderate

Marcucci G et al[124]	2010	Single cohort	Y	N/A	Non-industry	Low-moderate
Wagner K et al[21]	2010	Single cohort	Y	N/A	Non-industry	Low-moderate
Damm F et al[64]	2011	Y (275:293)	Y	none required	Non-industry	Low
Ho PA et al[20]	2010	Single cohort	Y	N/A	Non-industry	Low-moderate
Metzeler K et al[4]	2011	N (104:323)	Y	Y	Non-industry	Low
Becker H et al[13]	2010	Single cohort	Y	N/A	Non-industry	Low-moderate
Virappane et al[40]	2008	Single cohort	Y	N/A	Non-industry	Low-moderate
Marcucci G et al[61]	2008	Single cohort	Y	N/A	Non-industry	Low-moderate
Paschka P et al[63]	2008	Single cohort	Y	N/A	Non-industry	Low-moderate
Schwind S et al[70]	2010	Y (187:122)	Y	Y	Non-industry	Low
Hollink IH et al[87]	2009	Single cohort	Y	N/A	Non-industry	Low-moderate
Gaidzik VI et al[38]	2009	Single cohort	Y	N/A	Non-industry	Low-moderate
Damm F et al[59]	2010	N (249:50)	Y	No	NR	Moderate

Becker H et al[65]	20	Single cohort	Y	N/A	Non-industry	Low-moderate
Renneville A et al[117]	2012	Single cohort	Y	N/A	Non-industry	Low-moderate

Supplemental Table 17: Quality Assessment for Statement 17

Author	Year	Was there balance between treatment/assessment groups?	Reporting of baseline characteristics (and any diffs)	Reporting of any adjustment when diffs were present	Funding	Overall risk of bias assessment
Pollard JA et al[24]	2010	Single cohort	Yes (Y)	Not applicable (N/A)	Non-industry	Low-moderate
Markova J et al[33]	2009	Single cohort	Y	N/A	Non-industry	Low-moderate

Supplemental Table 18: Quality Assessment for Statement 19

Prospective Cohort Studies (N=2)						
Author	**Year**	**Was there balance between treatment/assessment groups?**	**Reporting of baseline characteristics (and any differences detected between groups)**	**Reporting of any adjustment when differences were present**	**Funding source**	**Overall risk of bias assessment**
Gaidzik VI et al[3]	2011	Single cohort	Yes (Y)	Not applicable (N/A)	Non-industry	Low-moderate
Pollard JA et al[24]	2010	Single cohort	Y	N/A	Non-industry	Low-moderate

Supplemental Table 19: Quality Assessment for Statement 20

Author	Year	Was there balance between treatment/assessment groups?	Reporting of baseline characteristics (and any differences detected between groups)	Reporting of any adjustment when differences were present	Funding source	Overall risk of bias assessment
Non-randomized Controlled Trials (N=2)						
Damm F et al[116]	2012	Single cohort	Not reported (NR)	NR	Non-industry	Low-moderate
Schwind S et al[119]	2011	Single cohort	NR	NR	Non-industry	Low-moderate
Prospective Cohort Studies (N=8)						
Alvarez S et al[68]	2010	No (N) (116:6)	Yes (Y)	NR	Non-industry	Low-moderate
Langer C et al[45]	2008	Single cohort	Y	Not applicable (N/A)	Non-industry	Low
Schwind S et al[70]	2010	Y (187:122)	Y	Y	Non-industry	Low
Marcucci G et al[75]	2008	Y (65:55)	Y	None required	Non-industry	Low
Schwind S et al[103]	2010	Single cohort	Y	N/A	Non-industry	Low-moderate
Chuang MK et al[120]	2015	N (138:57)	Y	N	Non-industry	Low
Metzeler K et al[4]	2011	N (104:323)	Y	Y	Non-industry	Low
Roman-Gomez J et	2009	Single cohort	Y	N/A	Non-industry	Low-moderate

al[101]						

Supplemental Table 20: Quality Assessment for Statement 23

Author	Year	Based on systematic review	Reproducible Methods	Quality assessment of included studies	Planned pooling stated a priori	Limitations of the study	Funding source	Overall risk of bias assessment
Meta-analyses (N=2)								
Krauter J et al[37]	2009	No (N)	Yes (Y)	Not reported (NR)	Y	NR	Non-industry	Low-moderate
Schaich M et al[51]	2007	N	Y	NR	Y	NR	Non-industry	Low-moderate
Randomized Controlled Trials (N=1)								
Vance G et al[109]	2007	NR, but was randomized	NR	NR	NR	NR	Non-industry	Moderate-high

Author	Year	Was there balance between treatment/assessment groups?	Reporting of baseline characteristics (and any differences detected between groups)	Reporting of any adjustment when differences were present	Funding source	Overall risk of bias assessment
Non-randomized Controlled Trials (N=2)						
Moorman AV et al[18]	2010	NR	NR	NR	Non-industry	Moderate
Arico M et al[47]	2008	N (115:1385:244:1744)	Y	NR	Non-industry	Low-moderate
Prospective Cohort Studies (N=23)						

Metzeler K et al[4]	2011	N (104:323)	Y	Y	Non-industry	Low
Montesinos P et al[6]	2011	N (72:579)	Y	Y	Non-industry	Low
Kayser S et al[8]	2011	N (200:2653)	Y	NR	Non-industry	Low-moderate
Clappier E et al[9]	2010	Single cohort	Y	Not applicable (N/A)	Non-industry	Low-moderate
Rollig C et al[14]	2010	Single cohort	Y	N/A	NR	Moderate
Grimwade D et al[19]	2010	Single cohort	N	N/A	Non-industry	Low-moderate
Basso G et al[31]	2009	Single cohort	Y	N/A	Non-industry	Low-moderate
Scrideli CA et al[36]	2009	Single cohort	Y	N/A	NR	Moderate
Lo-Coco F et al[42]	2008	Single cohort	N	N/A	Non-industry	Low-moderate
Forestier E et al[127]	2008	Single cohort	Y	N/A	Non-industry	Low-moderate
Wandt H et al[48]	2008	Single cohort	Y	N/A	Non-industry	Low-moderate
Barbaric D et al[54]	2007	N (1702:189)	Y	NR	Non-industry	Low-moderate
Mrózek K et al[57]	2008	N (4991:934)	N	NR	Non-industry	Moderate
Pabst T et	2009	Single cohort	N	N/A	Non-industry	Low-moderate

al[60]						
Marcucci G et al[61]	2008	Single cohort	Y	N/A	Non-industry	Low-moderate
Abbas S et al[71]	2010	Single cohort	Y	N/A	Non-industry	Low-moderate
Whitman SP et al[76]	2008	N (123:16)	Y	Y	Non-industry	Low
Chiaretti S et al[80]	2007	Single cohort	N	N/A	Non-industry	Moderate
Fischer L et al[83]	2009	Single cohort	Y	N/A	NR	Moderate
Medeiros BC et al[94]	2010	Single cohort	Y	N/A	NR	Moderate
Milani L et al[96]	2010	Single cohort	Y	N/A	Non-industry	Low-moderate
Moorman AV et al[108]	2007	Single cohort	Y	N/A	Non-industry	Low-moderate
Patel JL et al[111]	2012	Single cohort	N	N/A	Non-industry	Low-moderate

Supplemental Table 21: Quality Assessment for Statement 24

Author	Year	Was there balance between treatment/assessment groups?	Reporting of baseline characteristics (and any differences detected between groups)	Reporting of any adjustment when differences were present	Funding source	Overall risk of bias assessment

Barbaric D et al[54]	2007	No (N) (1702:189)	Yes (Y)	Not reported (NR)	Non-industry	Low-moderate
Mrózek K et al[57]	2008	N (4991:934)	N	NR	Non-industry	Moderate

Supplemental Table 22: Quality Assessment for Statement 27

Author	Year	Was there balance between treatment/assessment groups	Reporting of baseline characteristics (and any diffs)	Reporting of any adjustment when diffs were present	Funding	Overall risk of bias assessment
Prospective Cohort Studies (N=40)						
Metzeler K et al[4]	2011	No (N) (104:323)	Yes (Y)	Y	Non-industry	Low
Kayser S et al[91]	2009	Single cohort	Y	Not applicable (N/A)	Non-industry	Low-moderate
Meshinchi S et al[95]	2008	Single cohort	N	N/A	Non-industry	Low-moderate
Marks DI et al[30]	2009	Single cohort	Y	N/A	Non-industry	Low-moderate
Taskesen E et al[5]	2011	N (1031:60:91)	Y	Y	Non-industry	Low
Gaidzik VI et al[3]	2011	Single cohort	Y	N/A	Non-industry	Low-moderate
Clappier E et al[9]	2010	Single cohort	Y	N/A	Non-industry	Low-moderate

Kuhnl A et al[23]	2010	Single cohort	Y	N/A	Non-industry	Low-moderate
Pollard JA et al[24]	2010	Single cohort	Y	N/A	Non-industry	Low-moderate
Santamaria C et al[39]	2008	Single cohort	Not reported (NR)	N/A	Non-industry	Low-moderate
de Jonge HJ et al[72]	2010	Single cohort	Y	N/A	Non-industry	Low-moderate
Santamaria C et al[73]	2010	Single cohort	Y	N/A	Non-industry	Low-moderate
Csinady E et al[81]	2009	N (105:180)	N	NR	Non-industry	Moderate
Falini B et al[82]	2010	Single cohort	N	N/A	Non-industry	Moderate
Heesch S et al[86]	2010	Single cohort	Y	N/A	Non-industry	Low-moderate
Jiao B et al[89]	2009	Single cohort	Y	N/A	Non-industry	Low-moderate
Moorman AV et al[97]	2007	Single cohort	Y	N/A	Non-industry	Low-moderate
Ongaro A et al[99]	2009	Single cohort	Y	N/A	Non-industry	Low-moderate
Zachariadis V et al[107]	2011	Single cohort	Y	N/A	Non-industry	Low-moderate
Kuiper RP et al[69]	2010	N (102:29)	N	NR	Non-industry	Moderate

Kayser S et al[8]	2011	N (200:2653)	Y	NR	Non-industry	Low-moderate
Rollig C et al[14]	2010	Single cohort	Y	N/A	NR	Moderate
Harrison CJ et al[16]	2010	Single cohort	Y	N/A	Non-industry	Low-moderate
Buccisano F et al[12]	2010	Single cohort	Y	N/A	NR	Moderate
Becker H et al[13]	2010	Single cohort	Y	N/A	Non-industry	Low-moderate
Patel B et al[29]	2010	Single cohort	Y	N/A	Non-industry	Low-moderate
Schwind S et al[103]	2010	Single cohort	Y	N/A	Non-industry	Low-moderate
Lugthart S et al[11]	2010	N (94:52:19:123:2231)	Y	NR	Non-industry	Low-moderate
Maloney KW et al[15]	2010	N (80:2731)	Y	NR	Non-industry	Low-moderate
Salzer WL et al[26]	2010	Single cohort	Y	N/A	Non-industry	Low-moderate
Pui CH et al[28]	2010	N (325:188:165:247:53)	Y	N	Non-industry	Low-moderate
Cario G et al[79]	2010	Single cohort	Y	N/A	Non-industry	Low-moderate
Wheatley K et al[110]	2009	Single cohort	Y	N/A	NR	Moderate

Grimwade D et al[19]	2010	Single cohort	N	N/A	Non-industry	Low-moderate
Busse A et al[66]	2009	Single cohort	Y	N/A	NR	Moderate
Suela J et al[104]	2007	Single cohort	Y	N/A	Non-industry	Low-moderate
Moorman AV et al[108]	2007	Single cohort	Y	N/A	Non-industry	Low-moderate
Gonen M et al[113]	2012	Single cohort	Y	N/A	partial industry	Moderate
Moorman AV et al[114]	2012	Single cohort	N	N/A	Non-industry	Low-moderate
Patel JL et al[111]	2012	Single cohort	N	N/A	Non-industry	Low-moderate

Supplemental Table 23: Grades for Strength of Evidence

Designation	Description	Quality of Evidence
Convincing	High confidence that available evidence reflects true effect. Further research is very unlikely to change the confidence in the estimate of effect.	High/Intermediate quality evidence
Adequate	Moderate confidence that available evidence reflects true effect. Further research is likely to have an important impact on the confidence in	Intermediate/Low quality of evidence

	estimate of effect and may change the estimate.	
Inadequate	Little confidence that available evidence reflects true effect. Further research is very likely to have an important impact on the confidence in the estimate of effect and is likely to change the estimate.	Low/Insufficient evidence and expert panel uses formal consensus process to reach recommendation
Insufficient	Evidence is insufficient to discern net effect. Any estimate of effect is very uncertain.	Insufficient evidence and expert panel uses formal consensus process to reach recommendation

Adapted from J Clin Epidemiol, 64(4), Balshem H, Helfand M, Schunemann HJ, et l. GRADE guidelines: 3. Rating the quality of evidence, p. 401-406, copyright 2011, with permission from Elsevier.[128]

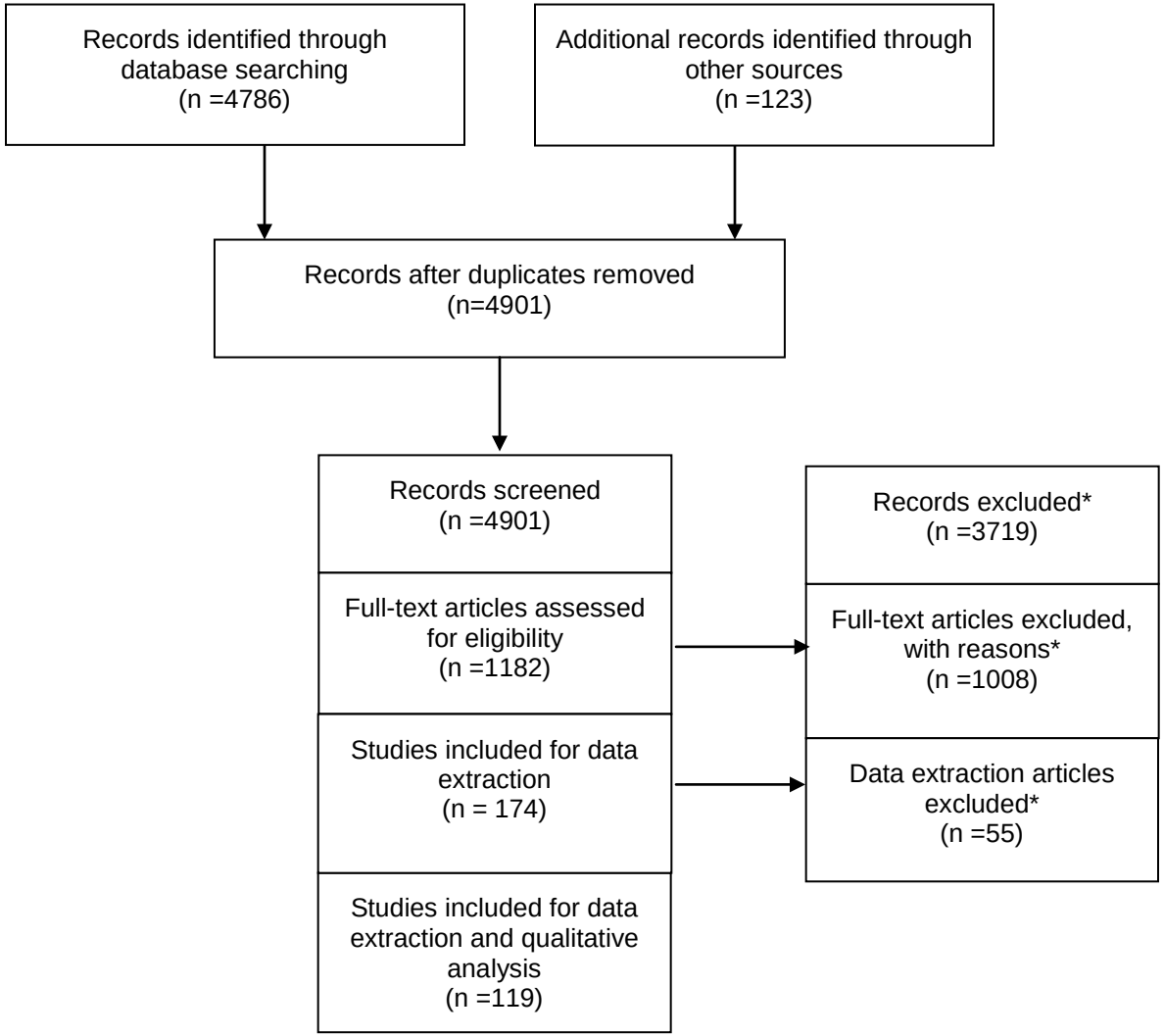
Supplemental Table 24: Grades for Strength of Recommendations

Designation	Recommendation	Rationale
Strong Recommendation	Recommend For or Against a particular practice (Can include must or should)	Supported by convincing (high) or adequate (intermediate) quality of evidence and clear benefit that outweighs any harms
Recommendation	Recommend For or Against a particular practice (Can include should or may)	Some limitations in quality of evidence (adequate [intermediate]), balance of benefits and harms, values, or costs but panel concludes that there is sufficient

		evidence to inform a recommendation
Expert Consensus Opinion	Recommend For or Against a particular practice (Can include should or may)	Serious limitations in quality of evidence (inadequate [low] or insufficient), balance of benefits and harms, values or costs, but panel consensus is that a statement is necessary
No Recommendation	No Recommendation For or Against a particular practice	Insufficient evidence to provide a recommendation, balance of benefits and harms, values or costs

Derived from Andrews et al.[129]

Supplemental Figure 1: Literature Review Flow Diagram **



*Excluded based on expert opinion, did not meet minimum quality standards, presented incomplete data or data that were not in useable formats
**Adapted from Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med.* 2009;6:e1000097.[130]

Supplemental Figure 2: Ovid Search Strings

001 leukemia, lymphoid/ or leukemia, biphenotypic, acute/ or precursor cell lymphoblastic leukemia-lymphoma/ or precursor b-cell lymphoblastic leukemia-lymphoma/ or precursor t-cell lymphoblastic leukemia-lymphoma/ or leukemia, myeloid, acute/ or leukemia, basophilic, acute/ or leukemia, eosinophilic, acute/ or leukemia, erythroblastic, acute/ or leukemia, megakaryoblastic, acute/ or leukemia, monocytic, acute/ or leukemia, promyelocytic, acute/ or leukemia, myelomonocytic, acute/ or sarcoma, myeloid/
002 "Anemia, Refractory, with Excess of Blasts"/
003 Leukemia, Myeloid/
004 Acute Disease/
005 exp Hematologic Neoplasms/
006 exp Dendritic Cells/
007 1 or 2 or 3 or 4 or 5 or 6
008 (Leuk?emia adj6 acute).tw.
009 (Sarcoma adj3 myeloid).tw.
010 (mixed phenotype adj3 leukemia).tw.
011 bilineal leukemia.tw.
012 "Secondary leukemia".tw.
013 "Therapy-related leukemia".tw.
014 "Granulocytic sarcoma".tw.
015 "blastic plasmacytoid dendritic cell tumor".tw.
016 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15
017 7 or 16
018 Diagnosis/
019 exp Diagnostic Errors/
020 "Diagnostic Techniques and Procedures"/
021 Neoplasm Staging/
022 exp Early Diagnosis/
023 Specimen Handling/
024 Prognosis/
025 Reproducibility of Results/
026 exp "Sensitivity and Specificity"/
027 clinical laboratory techniques/ or cytogenetic analysis/ or histological techniques/ or exp histocytochemistry/ or exp histocytological preparation techniques/ or molecular diagnostic techniques/
028 exp Polymerase Chain Reaction/
029 exp In Situ Hybridization, Fluorescence/
030 exp Immunohistochemistry/
031 Immunophenotyping/
032 Flow Cytometry/
033 exp Histocytochemistry/
034 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32 or 33
035 prognos:.tw.
036 Immunocytochemistry.tw.
037 35 or 36
038 34 or 37
039 17 and 38
040 limit 39 to english language
041 animal/
042 human/
043 human/ not (41 and 42)
044 40 and 43
045 limit 44 to (clinical trial, all or controlled clinical trial or meta analysis or multicenter study or randomized controlled trial or systematic reviews)
046 cohort stud:.mp.
047 prospective stud:.mp.

048 46 or 47
049 44 and 48
050 45 or 49
051 Case Reports.pt.
052 50 not 51

REFERENCES

1. Verkerk K, Van Veenendaal H, Severens JL, Hendriks EJ, Burgers JS. Considered judgement in evidence-based guideline development. *Int J Qual Health Care*. 2006;18(5):365-369.
2. IOM (Institute of Medicine). *Clinical Practice Guidelines We Can Trust*. Washington, DC: National Academies Press; 2011.
3. Gaidzik VI, Bullinger L, Schlenk RF, et al. RUNX1 mutations in acute myeloid leukemia: results from a comprehensive genetic and clinical analysis from the AML study group. *J Clin Oncol*. 2011;29(10):1364-1372.
4. Metzeler KH, Maharry K, Radmacher MD, et al. TET2 mutations improve the new European LeukemiaNet risk classification of acute myeloid leukemia: a Cancer and Leukemia Group B study. *J Clin Oncol*. 2011;29(10):1373-1381.
5. Taskesen E, Bullinger L, Corbacioglu A, et al. Prognostic impact, concurrent genetic mutations, and gene expression features of AML with CEBPA mutations in a cohort of 1182 cytogenetically normal AML patients: further evidence for CEBPA double mutant AML as a distinctive disease entity. *Blood*. 2011;117(8):2469-2475.
6. Montesinos P, Rayon C, Vellenga E, et al. Clinical significance of CD56 expression in patients with acute promyelocytic leukemia treated with all-trans retinoic acid and anthracycline-based regimens. *Blood*. 2011;117(6):1799-1805.
7. Stolzel F, Pfirrmann M, Aulitzky WE, et al. Risk stratification using a new prognostic score for patients with secondary acute myeloid leukemia: results of the prospective AML96 trial. *Leukemia*. 2011;25(3):420-428.
8. Kayser S, Dohner K, Krauter J, et al. The impact of therapy-related acute myeloid leukemia (AML) on outcome in 2853 adult patients with newly diagnosed AML. *Blood*. 2011;117(7):2137-2145.
9. Clappier E, Collette S, Grardel N, et al. NOTCH1 and FBXW7 mutations have a favorable impact on early response to treatment, but not on outcome, in children with T-cell acute lymphoblastic leukemia (T-ALL) treated on EORTC trials 58881 and 58951. *Leukemia*. 2010;24(12):2023-2031.
10. Tallman MS, Kim HT, Montesinos P, et al. Does microgranular variant morphology of acute promyelocytic leukemia independently predict a less favorable outcome compared with classical M3 APL? A joint study of the North American Intergroup and the PETHEMA Group. *Blood*. 2010;116(25):5650-5659.
11. Lugthart S, Groschel S, Beverloo HB, et al. Clinical, molecular, and prognostic significance of WHO type inv(3)(q21q26.2)/t(3;3)(q21;q26.2) and various other 3q abnormalities in acute myeloid leukemia. *J Clin Oncol*. 2010;28(24):3890-3898.
12. 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.

13. Becker H, Marcucci G, Maharry K, et al. Mutations of the Wilms tumor 1 gene (WT1) in older patients with primary cytogenetically normal acute myeloid leukemia: a Cancer and Leukemia Group B study. *Blood*. 2010;116(5):788-792.

14. Rollig C, Thiede C, Gramatzki M, et al. A novel prognostic model in elderly patients with acute myeloid leukemia: results of 909 patients entered into the prospective AML96 trial. *Blood*. 2010;116(6):971-978.

15. Maloney KW, Carroll WL, Carroll AJ, et al. Down syndrome childhood acute lymphoblastic leukemia has a unique spectrum of sentinel cytogenetic lesions that influences treatment outcome: a report from the Children's Oncology Group. *Blood*. 2010;116(7):1045-1050.

16. Harrison CJ, Hills RK, Moorman AV, et al. Cytogenetics of childhood acute myeloid leukemia: United Kingdom Medical Research Council Treatment trials AML 10 and 12. *J Clin Oncol*. 2010;28(16):2674-2681.

17. Ho PA, Zeng R, Alonzo TA, et al. Prevalence and prognostic implications of WT1 mutations in pediatric acute myeloid leukemia (AML): a report from the Children's Oncology Group. *Blood*. 2010;116(5):702-710.

18. Moorman AV, Ensor HM, Richards SM, et al. Prognostic effect of chromosomal abnormalities in childhood B-cell precursor acute lymphoblastic leukaemia: results from the UK Medical Research Council ALL97/99 randomised trial. *Lancet Oncol*. 2010;11(5):429-438.

19. Grimwade D, Hills RK, Moorman AV, et al. Refinement of cytogenetic classification in acute myeloid leukemia: determination of prognostic significance of rare recurring chromosomal abnormalities among 5876 younger adult patients treated in the United Kingdom Medical Research Council trials. *Blood*. 2010;116(3):354-365.

20. Ho PA, Alonzo TA, Kopecky KJ, et al. Molecular alterations of the IDH1 gene in AML: a Children's Oncology Group and Southwest Oncology Group study. *Leukemia*. 2010;24(5):909-913.

21. Wagner K, Damm F, Gohring G, et al. Impact of IDH1 R132 mutations and an IDH1 single nucleotide polymorphism in cytogenetically normal acute myeloid leukemia: SNP rs11554137 is an adverse prognostic factor. *J Clin Oncol*. 2010;28(14):2356-2364.

22. Groschel S, Lugthart S, Schlenk RF, et al. High EVI1 expression predicts outcome in younger adult patients with acute myeloid leukemia and is associated with distinct cytogenetic abnormalities. *J Clin Oncol*. 2010;28(12):2101-2107.

23. Kuhnl A, Gokbuget N, Stroux A, et al. High BAALC expression predicts chemoresistance in adult B-precursor acute lymphoblastic leukemia. *Blood*. 2010;115(18):3737-3744.

24. Pollard JA, Alonzo TA, Gerbing RB, et al. Prevalence and prognostic significance of KIT mutations in pediatric patients with core binding factor AML enrolled on serial pediatric cooperative trials for de novo AML. *Blood*. 2010;115(12):2372-2379.

25. Escherich G, Horstmann MA, Zimmermann M, Janka-Schaub GE. Cooperative study group for childhood acute lymphoblastic leukaemia (COALL): long-term results of trials 82,85,89,92 and 97. *Leukemia*. 2010;24(2):298-308.

26. Salzer WL, Devidas M, Carroll WL, et al. Long-term results of the pediatric oncology group studies for childhood acute lymphoblastic leukemia 1984-2001: a report from the Children's Oncology Group. *Leukemia*. 2010;24(2):355-370.

27. Schmiegelow K, Forestier E, Hellebostad M, et al. Long-term results of NOPHO ALL-92 and ALL-2000 studies of childhood acute lymphoblastic leukemia. *Leukemia*. 2010;24(2):345-354.

28. Pui CH, Pei D, Sandlund JT, et al. Long-term results of St Jude Total Therapy Studies 11, 12, 13A, 13B, and 14 for childhood acute lymphoblastic leukemia. *Leukemia*. 2010;24(2):371-382.

29. Patel B, Rai L, Buck G, et al. Minimal residual disease is a significant predictor of treatment failure in non T-lineage adult acute lymphoblastic leukaemia: final results of the international trial UKALL XII/ECOG2993. *Br J Haematol*. 2010;148(1):80-89.

30. Marks DI, Paietta EM, Moorman AV, et al. T-cell acute lymphoblastic leukemia in adults: clinical features, immunophenotype, cytogenetics, and outcome from the large randomized prospective trial (UKALL XII/ECOG 2993). *Blood*. 2009;114(25):5136-5145.

31. Basso G, Veltroni M, Valsecchi MG, et al. Risk of relapse of childhood acute lymphoblastic leukemia is predicted by flow cytometric measurement of residual disease on day 15 bone marrow. *J Clin Oncol*. 2009;27(31):5168-5174.

32. Metzeler KH, Dufour A, Benthaus T, et al. ERG expression is an independent prognostic factor and allows refined risk stratification in cytogenetically normal acute myeloid leukemia: a comprehensive analysis of ERG, MN1, and BAALC transcript levels using oligonucleotide microarrays. *J Clin Oncol*. 2009;27(30):5031-5038.

33. Markova J, Trnkova Z, Michkova P, et al. Monitoring of minimal residual disease in patients with core binding factor acute myeloid leukemia and the impact of C-KIT, FLT3, and JAK2 mutations on clinical outcome. *Leuk Lymphoma*. 2009;50(9):1448-1460.

34. Familiades J, Bousquet M, Lafage-Pochitaloff M, et al. PAX5 mutations occur frequently in adult B-cell progenitor acute lymphoblastic leukemia and PAX5 haploinsufficiency is associated with BCR-ABL1 and TCF3-PBX1 fusion genes: a GRAALL study. *Leukemia*. 2009;23(11):1989-1998.

35. Karrman K, Forestier E, Heyman M, et al. Clinical and cytogenetic features of a population-based consecutive series of 285 pediatric T-cell acute lymphoblastic leukemias: rare T-cell receptor gene rearrangements are associated with poor outcome. *Genes Chromosomes Cancer*. 2009;48(9):795-805.

36. Scrideli CA, Assumpcao JG, Ganazza MA, et al. A simplified minimal residual disease polymerase chain reaction method at early treatment points can stratify children with acute lymphoblastic leukemia into good and poor outcome groups. *Haematologica*. 2009;94(6):781-789.

37. Krauter J, Wagner K, Schafer I, et al. Prognostic factors in adult patients up to 60 years old with acute myeloid leukemia and translocations of chromosome band 11q23: individual patient data-based meta-analysis of the German Acute Myeloid Leukemia Intergroup. *J Clin Oncol*. 2009;27(18):3000-3006.

38. Gaidzik VI, Schlenk RF, Moschny S, et al. Prognostic impact of WT1 mutations in cytogenetically normal acute myeloid leukemia: a study of the German-Austrian AML Study Group. *Blood*. 2009;113(19):4505-4511.

39. Santamaria C, Chillon MC, Garcia-Sanz R, et al. The relevance of preferentially expressed antigen of melanoma (PRAME) as a marker of disease activity and prognosis in acute promyelocytic leukemia. *Haematologica*. 2008;93(12):1797-1805.

40. Virappane P, Gale R, Hills R, et al. Mutation of the Wilms' tumor 1 gene is a poor prognostic factor associated with chemotherapy resistance in normal karyotype acute myeloid

leukemia: the United Kingdom Medical Research Council Adult Leukaemia Working Party. *J Clin Oncol*. 2008;26(33):5429-5435.

41. Pulsoni A, Iacobelli S, Bernardi M, et al. M4 acute myeloid leukemia: the role of eosinophilia and cytogenetics in treatment response and survival. The GIMEMA experience. *Haematologica*. 2008;93(7):1025-1032.

42. Lo-Coco F, Cuneo A, Pane F, et al. Prognostic impact of genetic characterization in the GIMEMA LAM99P multicenter study for newly diagnosed acute myeloid leukemia. *Haematologica*. 2008;93(7):1017-1024.

43. Holowiecki J, Krawczyk-Kulis M, Giebel S, et al. Status of minimal residual disease after induction predicts outcome in both standard and high-risk Ph-negative adult acute lymphoblastic leukaemia: the Polish Adult Leukemia Group ALL 4-2002 MRD Study. *Br J Haematol*. 2008;142(2):227-237.

44. Rubnitz JE, Wichlan D, Devidas M, et al. Prospective analysis of TEL gene rearrangements in childhood acute lymphoblastic leukemia: a Children's Oncology Group study. *J Clin Oncol*. 2008;26(13):2186-2191.

45. Langer C, Radmacher MD, Ruppert AS, et al. High BAALC expression associates with other molecular prognostic markers, poor outcome, and a distinct gene-expression signature in cytogenetically normal patients younger than 60 years with acute myeloid leukemia: a Cancer and Leukemia Group B (CALGB) study. *Blood*. 2008;111(11):5371-5379.

46. Oudot C, Auclerc MF, Levy V, et al. Prognostic factors for leukemic induction failure in children with acute lymphoblastic leukemia and outcome after salvage therapy: the FRALLE 93 study. *J Clin Oncol*. 2008;26(9):1496-1503.

47. Arico M, Valsecchi MG, Rizzari C, et al. Long-term results of the AIEOP-ALL-95 Trial for Childhood Acute Lymphoblastic Leukemia: insight on the prognostic value of DNA index in the framework of Berlin-Frankfurt-Muenster based chemotherapy. *J Clin Oncol*. 2008;26(2):283-289.

48. Wandt H, Schakel U, Kroschinsky F, et al. MLD according to the WHO classification in AML has no correlation with age and no independent prognostic relevance as analyzed in 1766 patients. *Blood*. 2008;111(4):1855-1861.

49. Lange BJ, Smith FO, Feusner J, et al. Outcomes in CCG-2961, a children's oncology group phase 3 trial for untreated pediatric acute myeloid leukemia: a report from the Children's Oncology Group. *Blood*. 2008;111(3):1044-1053.

50. Gale RE, Green C, Allen C, et al. The impact of FLT3 internal tandem duplication mutant level, number, size, and interaction with NPM1 mutations in a large cohort of young adult patients with acute myeloid leukemia. *Blood*. 2008;111(5):2776-2784.

51. Schaich M, Schlenk RF, Al-Ali HK, et al. Prognosis of acute myeloid leukemia patients up to 60 years of age exhibiting trisomy 8 within a non-complex karyotype: individual patient data-based meta-analysis of the German Acute Myeloid Leukemia Intergroup. *Haematologica*. 2007;92(6):763-770.

52. Zhou J, Goldwasser MA, Li A, et al. Quantitative analysis of minimal residual disease predicts relapse in children with B-lineage acute lymphoblastic leukemia in DFCI ALL Consortium Protocol 95-01. *Blood*. 2007;110(5):1607-1611.

53. Yanada M, Matsushita T, Asou N, et al. Severe hemorrhagic complications during remission induction therapy for acute promyelocytic leukemia: incidence, risk factors, and influence on outcome. *Eur J Haematol*. 2007;78(3):213-219.

54. Barbaric D, Alonzo TA, Gerbing RB, et al. Minimally differentiated acute myeloid leukemia (FAB AML-M0) is associated with an adverse outcome in children: a report from the Children's Oncology Group, studies CCG-2891 and CCG-2961. *Blood*. 2007;109(6):2314-2321.

55. Roman-Gomez J, Cordeu L, Agirre X, et al. Epigenetic regulation of Wnt-signaling pathway in acute lymphoblastic leukemia. *Blood*. 2007;109(8):3462-3469.

56. Asnafi V, Buzyn A, Le Noir S, et al. NOTCH1/FBXW7 mutation identifies a large subgroup with favorable outcome in adult T-cell acute lymphoblastic leukemia (T-ALL): a Group for Research on Adult Acute Lymphoblastic Leukemia (GRAALL) study. *Blood*. 2009;113(17):3918-3924.

57. Mrózek K, Carroll AJ, Maharry K, et al. Central review of cytogenetics is necessary for cooperative group correlative and clinical studies of adult acute leukemia: the Cancer and Leukemia Group B experience. *Int J Oncol*. 2008;33(2):239-244.

58. Dufour A, Schneider F, Metzeler KH, et al. Acute myeloid leukemia with biallelic CEBPA gene mutations and normal karyotype represents a distinct genetic entity associated with a favorable clinical outcome. *J Clin Oncol*. 2010;28(4):570-577.

59. Damm F, Heuser M, Morgan M, et al. Single nucleotide polymorphism in the mutational hotspot of WT1 predicts a favorable outcome in patients with cytogenetically normal acute myeloid leukemia. *J Clin Oncol*. 2010;28(4):578-585.

60. Pabst T, Eyholzer M, Fos J, Mueller BU. Heterogeneity within AML with CEBPA mutations; only CEBPA double mutations, but not single CEBPA mutations are associated with favourable prognosis. *Br J Cancer*. 2009;100(8):1343-1346.

61. Marcucci G, Maharry K, Radmacher MD, et al. Prognostic significance of, and gene and microRNA expression signatures associated with, CEBPA mutations in cytogenetically normal acute myeloid leukemia with high-risk molecular features: a Cancer and Leukemia Group B Study. *J Clin Oncol*. 2008;26(31):5078-5087.

62. Schneider F, Hoster E, Unterhalt M, et al. NPM1 but not FLT3-ITD mutations predict early blast cell clearance and CR rate in patients with normal karyotype AML (NK-AML) or high-risk myelodysplastic syndrome (MDS). *Blood*. 2009;113(21):5250-5253.

63. Paschka P, Marcucci G, Ruppert AS, et al. Wilms' tumor 1 gene mutations independently predict poor outcome in adults with cytogenetically normal acute myeloid leukemia: a cancer and leukemia group B study. *J Clin Oncol*. 2008;26(28):4595-4602.

64. Damm F, Heuser M, Morgan M, et al. Integrative prognostic risk score in acute myeloid leukemia with normal karyotype. *Blood*. 2011;117(17):4561-4568.

65. Becker H, Marcucci G, Maharry K, et al. Favorable prognostic impact of NPM1 mutations in older patients with cytogenetically normal de novo acute myeloid leukemia and associated gene- and microRNA-expression signatures: a Cancer and Leukemia Group B study. *J Clin Oncol*. 2010;28(4):596-604.

66. Busse A, Gokbuget N, Siehl JM, et al. Wilms' tumor gene 1 (WT1) expression in subtypes of acute lymphoblastic leukemia (ALL) of adults and impact on clinical outcome. *Ann Hematol*. 2009;88(12):1199-1205.

67. Rubnitz JE, Razzouk BI, Lensing S, Pounds S, Pui C-H, Ribeiro RC. Prognostic factors and outcome of recurrence in childhood acute myeloid leukemia. *Cancer*. 2007;109(1):157-163.

68. Alvarez S, Suela J, Valencia A, et al. DNA methylation profiles and their relationship with cytogenetic status in adult acute myeloid leukemia. *PLoS One*. 2010;5(8):e12197. doi: 10.1371/journal.pone.012197

69. Kuiper RP, Waanders E, van der Velden VH, et al. IKZF1 deletions predict relapse in uniformly treated pediatric precursor B-ALL. *Leukemia*. 2010;24(7):1258-1264.

70. Schwind S, Maharry K, Radmacher MD, et al. Prognostic significance of expression of a single microRNA, miR-181a, in cytogenetically normal acute myeloid leukemia: a Cancer and Leukemia Group B study. *J Clin Oncol*. 2010;28(36):5257-5264.

71. Abbas S, Lugthart S, Kavelaars FG, et al. Acquired mutations in the genes encoding IDH1 and IDH2 both are recurrent aberrations in acute myeloid leukemia: prevalence and prognostic value. *Blood*. 2010;116(12):2122-2126.

72. de Jonge HJ, Valk PJ, Veeger NJ, et al. High VEGFC expression is associated with unique gene expression profiles and predicts adverse prognosis in pediatric and adult acute myeloid leukemia. *Blood*. 2010;116(10):1747-1754.

73. Santamaria C, Chillon MC, Garcia-Sanz R, et al. BAALC is an important predictor of refractoriness to chemotherapy and poor survival in intermediate-risk acute myeloid leukemia (AML). *Ann Hematol*. 2010;89(5):453-458.

74. Seifert H, Mohr B, Thiede C, et al. The prognostic impact of 17p (p53) deletion in 2272 adults with acute myeloid leukemia. *Leukemia*. 2009;23(4):656-663.

75. Marcucci G, Radmacher MD, Maharry K, et al. MicroRNA expression in cytogenetically normal acute myeloid leukemia. *N Engl J Med*. 2008;358(18):1919-1928.

76. Whitman SP, Ruppert AS, Radmacher MD, et al. FLT3 D835/I836 mutations are associated with poor disease-free survival and a distinct gene-expression signature among younger adults with de novo cytogenetically normal acute myeloid leukemia lacking FLT3 internal tandem duplications. *Blood*. 2008;111(3):1552-1559.

77. Baldus CD, Thibaut J, Goekbuget N, et al. Prognostic implications of NOTCH1 and FBXW7 mutations in adult acute T-lymphoblastic leukemia. *Haematologica*. 2009;94(10):1383-1390.

78. Breems DA, Van Putten WLJ, De Greef GE, et al. Monosomal karyotype in acute myeloid leukemia: a better indicator of poor prognosis than a complex karyotype. *J Clin Oncol*. 2008;26(29):4791-4797.

79. Cario G, Zimmermann M, Romey R, et al. Presence of the P2RY8-CRLF2 rearrangement is associated with a poor prognosis in non-high-risk precursor B-cell acute lymphoblastic leukemia in children treated according to the ALL-BFM 2000 protocol. *Blood*. 2010;115(26):5393-5397.

80. Chiaretti S, Tavolaro S, Ghia EM, et al. Characterization of ABL1 expression in adult T-cell acute lymphoblastic leukemia by oligonucleotide array analysis. *Haematologica*. 2007;92(5):619-626.

81. Csinady E, van der Velden VHJ, Joas R, et al. Chromosome 14 copy number-dependent IGH gene rearrangement patterns in high hyperdiploid childhood B-cell precursor ALL: implications for leukemia biology and minimal residual disease analysis. *Leukemia*. 2009;23(5):870-876.

82. Falini B, Macijewski K, Weiss T, et al. Multilineage dysplasia has no impact on biologic, clinicopathologic, and prognostic features of AML with mutated nucleophosmin (NPM1). *Blood*. 2010;115(18):3776-3786.

83. Fischer L, Gokbuget N, Schwartz S, et al. CD56 expression in T-cell acute lymphoblastic leukemia is associated with non-thymic phenotype and resistance to induction therapy but no inferior survival after risk-adapted therapy. *Haematologica*. 2009;94(2):224-229.

84. Flex E, Petrangeli V, Stella L, et al. Somatically acquired JAK1 mutations in adult acute lymphoblastic leukemia. *J Exp Med*. 2008;205(4):751-758.

85. Harvey RC, Mullighan CG, Wang X, et al. Identification of novel cluster groups in pediatric high-risk B-precursor acute lymphoblastic leukemia with gene expression profiling: correlation with genome-wide DNA copy number alterations, clinical characteristics, and outcome. *Blood*. 2010;116(23):4874-4884.

86. Heesch S, Goekbuget N, Stroux A, et al. Prognostic implications of mutations and expression of the Wilms tumor 1 (WT1) gene in adult acute T-lymphoblastic leukemia. *Haematologica*. 2010;95(6):942-949.

87. Hollink IH, van den Heuvel-Eibrink MM, Zimmermann M, et al. Clinical relevance of Wilms tumor 1 gene mutations in childhood acute myeloid leukemia. *Blood*. 2009;113(23):5951-5960.

88. Jeha S, Pei D, Raimondi SC, et al. Increased risk for CNS relapse in pre-B cell leukemia with the t(1;19)/TCF3-PBX1. *Leukemia*. 2009;23(8):1406-1409.

89. Jiao B, Wu CF, Liang Y, et al. AML1-ETO9a is correlated with C-KIT overexpression/mutations and indicates poor disease outcome in t(8;21) acute myeloid leukemia-M2. *Leukemia*. 2009;23(9):1598-1604.

90. Johnston DL, Alonzo TA, Gerbing RB, Lange BJ, Woods WG. The presence of central nervous system disease at diagnosis in pediatric acute myeloid leukemia does not affect survival: a Children's Oncology Group study. *Pediatr Blood Cancer*. 2010;55(3):414-420.

91. Kayser S, Schlenk RF, Londono MC, et al. Insertion of FLT3 internal tandem duplication in the tyrosine kinase domain-1 is associated with resistance to chemotherapy and inferior outcome. *Blood*. 2009;114(12):2386-2392.

92. Kox C, Zimmermann M, Stanulla M, et al. The favorable effect of activating NOTCH1 receptor mutations on long-term outcome in T-ALL patients treated on the ALL-BFM 2000 protocol can be separated from FBXW7 loss of function. *Leukemia*. 2010;24(12):2005-2013.

93. Maurillo L, Buccisano F, Del Principe MI, et al. Toward optimization of postremission therapy for residual disease-positive patients with acute myeloid leukemia. *J Clin Oncol*. 2008;26(30):4944-4951.

94. Medeiros BC, Othus M, Fang M, Roulston D, Appelbaum FR. Prognostic impact of monosomal karyotype in young adult and elderly acute myeloid leukemia: the Southwest Oncology Group (SWOG) experience. *Blood*. 2010;116(13):2224-2228.

95. Meshinchi S, Stirewalt DL, Alonzo TA, et al. Structural and numerical variation of FLT3/ITD in pediatric AML. *Blood*. 2008;111(10):4930-4933.

96. Milani L, Lundmark A, Kiialainen A, et al. DNA methylation for subtype classification and prediction of treatment outcome in patients with childhood acute lymphoblastic leukemia. *Blood*. 2010;115(6):1214-1225.

97. Moorman AV, Richards SM, Robinson HM, et al. Prognosis of children with acute lymphoblastic leukemia (ALL) and intrachromosomal amplification of chromosome 21 (iAMP21). *Blood*. 2007;109(6):2327-2330.

98. Mullighan CG, Su X, Zhang J, et al. Deletion of IKZF1 and prognosis in acute lymphoblastic leukemia. *N Engl J Med*. 2009;360(5):470-480.

99. Ongaro A, De Mattei M, Della Porta MG, et al. Gene polymorphisms in folate metabolizing enzymes in adult acute lymphoblastic leukemia: effects on methotrexate-related toxicity and survival. *Haematologica*. 2009;94(10):1391-1398.

100. Prebet T, Boissel N, Reutenauer S, et al. Acute myeloid leukemia with translocation (8;21) or inversion (16) in elderly patients treated with conventional chemotherapy: a collaborative study of the French CBF-AML intergroup. *J Clin Oncol*. 2009;27(28):4747-4753.

101. Roman-Gomez J, Agirre X, Jimenez-Velasco A, et al. Epigenetic regulation of microRNAs in acute lymphoblastic leukemia. *J Clin Oncol*. 2009;27(8):1316-1322.

102. Santamaria CM, Chillon MC, Garcia-Sanz R, et al. High FOXO3a expression is associated with a poorer prognosis in AML with normal cytogenetics. *Leuk Res*. 2009;33(12):1706-1709.

103. Schwind S, Marcucci G, Maharry K, et al. BAALC and ERG expression levels are associated with outcome and distinct gene and microRNA expression profiles in older patients with de novo cytogenetically normal acute myeloid leukemia: a Cancer and Leukemia Group B study. *Blood*. 2010;116(25):5660-5669.

104. Suela J, Alvarez S, Cifuentes F, et al. DNA profiling analysis of 100 consecutive de novo acute myeloid leukemia cases reveals patterns of genomic instability that affect all cytogenetic risk groups. *Leukemia*. 2007;21(6):1224-1231.

105. Tauchi H, Tomizawa D, Eguchi M, et al. Clinical features and outcome of MLL gene rearranged acute lymphoblastic leukemia in infants with additional chromosomal abnormalities other than 11q23 translocation. *Leuk Res*. 2008;32(10):1523-1529.

106. Waanders E, van der Velden VHJ, van der Schoot CE, et al. Integrated use of minimal residual disease classification and IKZF1 alteration status accurately predicts 79% of relapses in pediatric acute lymphoblastic leukemia. *Leukemia*. 2011;25(2):254-258.

107. Zachariadis V, Gauffin F, Kuchinskaya E, et al. The frequency and prognostic impact of dic(9;20)(p13.2;q11.2) in childhood B-cell precursor acute lymphoblastic leukemia: results from the NOPHO ALL-2000 trial. *Leukemia*. 2011;25(4):622-628.

108. Moorman AV, Harrison CJ, Buck GA, et al. Karyotype is an independent prognostic factor in adult acute lymphoblastic leukemia (ALL): analysis of cytogenetic data from patients treated on the Medical Research Council (MRC) UKALLXII/Eastern Cooperative Oncology Group (ECOG) 2993 trial. *Blood*. 2007;109(8):3189-3197.

109. Vance GH, Kim H, Hicks GA, et al. Utility of interphase FISH to stratify patients into cytogenetic risk categories at diagnosis of AML in an Eastern Cooperative Oncology Group (ECOG) clinical trial (E1900). *Leuk Res*. 2007;31(5):605-609.

110. Wheatley K, Brookes CL, Howman AJ, et al. Prognostic factor analysis of the survival of elderly patients with AML in the MRC AML11 and LRF AML14 trials. *Br J Haematol*. 2009;145(5):598-605.

111. Patel JL, Smith LM, Anderson J, et al. The immunophenotype of T-lymphoblastic lymphoma in children and adolescents: a Children's Oncology Group report. *Br J Haematol*. 2012;159(4):454-461.

112. Yin JAL, O'Brien MA, Hills RK, Daly SB, Wheatley K, Burnett AK. Minimal residual disease monitoring by quantitative RT-PCR in core binding factor AML allows risk stratification and predicts relapse: results of the United Kingdom MRC AML-15 trial. *Blood*. 2012;120(14):2826-2835.

113. Gonen M, Sun Z, Figueroa ME, et al. CD25 expression status improves prognostic risk classification in AML independent of established biomarkers: ECOG phase 3 trial, E1900. *Blood*. 2012;120(11):2297-2306.

114. Moorman AV, Schwab C, Ensor HM, et al. IGH@ translocations, CRLF2 deregulation, and microdeletions in adolescents and adults with acute lymphoblastic leukemia. *J Clin Oncol*. 2012;30(25):3100-3108.

115. Mendler JH, Maharry K, Radmacher MD, et al. RUNX1 mutations are associated with poor outcome in younger and older patients with cytogenetically normal acute myeloid leukemia and with distinct gene and MicroRNA expression signatures. *J Clin Oncol*. 2012;30(25):3109-3118.

116. Damm F, Wagner K, Gorlich K, et al. ID1 expression associates with other molecular markers and is not an independent prognostic factor in cytogenetically normal acute myeloid leukaemia. *Br J Haematol*. 2012;158(2):208-215.

117. Renneville A, Boissel N, Nibourel O, et al. Prognostic significance of DNA methyltransferase 3A mutations in cytogenetically normal acute myeloid leukemia: a study by the Acute Leukemia French Association. *Leukemia*. 2012;26(6):1247-1254.

118. Schneider F, Hoster E, Unterhalt M, et al. The FLT3ITD mRNA level has a high prognostic impact in NPM1 mutated, but not in NPM1 unmutated, AML with a normal karyotype. *Blood*. 2012;119(19):4383-4386.

119. Schwind S, Marcucci G, Kohlschmidt J, et al. Low expression of MN1 associates with better treatment response in older patients with de novo cytogenetically normal acute myeloid leukemia. *Blood*. 2011;118(15):4188-4198.

120. Chuang MK, Chiu YC, Chou WC, Hou HA, Chuang EY, Tien HF. A 3-microRNA scoring system for prognostication in de novo acute myeloid leukemia patients. *Leukemia*. 2015;29(5):1051-1059.

121. Schlenk RF, Kayser S, Bullinger L, et al. Differential impact of allelic ratio and insertion site in FLT3-ITD-positive AML with respect to allogeneic transplantation. *Blood*. 2014;124(23):3441-3449.

122. Den Boer ML, van Slegtenhorst M, De Menezes RX, et al. A subtype of childhood acute lymphoblastic leukaemia with poor treatment outcome: a genome-wide classification study. *Lancet Oncol*. 2009;10(2):125-134.

123. Paschka P, Schlenk RF, Gaidzik VI, et al. IDH1 and IDH2 mutations are frequent genetic alterations in acute myeloid leukemia and confer adverse prognosis in cytogenetically normal acute myeloid leukemia with NPM1 mutation without FLT3 internal tandem duplication. *J Clin Oncol*. 2010;28(22):3636-3643.

124. Marcucci G, Maharry K, Wu YZ, et al. IDH1 and IDH2 gene mutations identify novel molecular subsets within de novo cytogenetically normal acute myeloid leukemia: a Cancer and Leukemia Group B study. *J Clin Oncol*. 2010;28(14):2348-2355.

125. Zhou K-G, Jiang L-J, Shang Z, Wang J, Huang L, Zhou J-F. Potential application of IDH1 and IDH2 mutations as prognostic indicators in non-promyelocytic acute myeloid leukemia: a meta-analysis. *Leuk Lymphoma*. 2012;53(12):2423-2429.

126. Nomdedeu J, Hoyos M, Carricondo M, et al. Adverse impact of IDH1 and IDH2 mutations in primary AML: experience of the Spanish CETLAM group. *Leuk Res*. 2012;36(8):990-997.

127. Forestier E, Heyman M, Andersen MK, et al. Outcome of ETV6/RUNX1-positive childhood acute lymphoblastic leukaemia in the NOPHO-ALL-1992 protocol: frequent late relapses but good overall survival. *Br J Haematol*. 2008;140(6):665-672.

128. Balshem H, Helfand M, Schunemann HJ, et al. GRADE guidelines, 3: rating the quality of evidence. *J Clin Epidemiol*. 2011;64(4):401-406.

129. Andrews J, Guyatt G, Oxman AD, et al. GRADE guidelines, 14: going from evidence to recommendations: the significance and presentation of recommendations. *J Clin Epidemiol*. 2013;66(7):719-725.

130. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med*. 2009;6(7):e1000097. doi: 10.1371/journal.pmed.1000097