Next-generation sequencing to detect measurable residual disease (MRD) at a threshold of $10^4$ (one cell in 10,000) as an alternative to standard testing (e.g., flow cytometry or polymerase chain reaction) in patients with acute lymphoblastic leukemia or multiple myeloma may be considered medically necessary (see Policy Guidelines below for details related to this statement).

Next-generation sequencing to detect measurable residual disease (MRD) at a threshold of below $10^4$ as a replacement test in patients with acute lymphoblastic leukemia or multiple myeloma is considered investigational.

Next-generation sequencing to detect MRD as a replacement test is considered investigational in all other situations.

Policy Guidelines

Standard testing methods for assessing measurable residual disease include flow cytometry and polymerase chain reaction (PCR). Their threshold of detection is usually set at $10^4$. Next Generation Sequencing (NGS) has been found to be accurate at even lower levels of detection of residual cancer cells. However, the clinical utility of that lower level of detection is unclear. It is unclear if such detection would result in some overtreatment along with possible complications. Therefore, testing for measurable residual disease below the $10^4$ level of detection is of uncertain benefit, and does not offer clear advantages at this time. Residual Disease using NGS may be medically necessary testing for the treatment of acute lymphoblastic leukemia and multiple myeloma cancers. Flow cytometry and PCR are also an effective evaluation for these diagnoses and medically necessary. When there are two medically necessary procedures for the evaluation or treatment of acute lymphoblastic leukemia or multiple myeloma cancer, Blue Shield will consider the relative cost of each and provide coverage for the procedure that is most cost effective. The other test will be denied as not cost effective, and therefore not medically necessary under the circumstances. NGS for measurable residual disease (e.g., ClonSEQ) is not cost effective compared with standard testing within established parameters and is therefore considered not medically necessary.

There is no specific code for next generation sequencing for measurable residual disease monitoring. ClonoSEQ® Minimal Residual Disease Test may be billed with the following unlisted codes:

- 81599: Unlisted multianalyte assay with algorithmic analysis
- 81479: Unlisted molecular pathology procedure

Description

Measurable residual disease (MRD), also known as minimal residual disease, refers to residual clonal cells in blood or bone marrow following treatment for hematologic malignancies. MRD is typically assessed by flow cytometry or polymerase chain reaction, which can detect one clonal cell in 100,000 cells. It is proposed that next-generation sequencing (NGS), which can detect one residual clonal sequence out of 1,000,000 cells, will improve health outcomes in patients who have been treated for hematologic malignancies.
Related Policies

- Hematopoietic Cell Transplantation for Acute Lymphoblastic Leukemia
- Hematopoietic Cell Transplantation for Acute Myeloid Leukemia
- Hematopoietic Cell Transplantation for Chronic Lymphocytic Leukemia Small Lymphocytic Lymphoma
- Hematopoietic Cell Transplantation for Chronic Myeloid Leukemia
- Hematopoietic Cell Transplantation for Hodgkin Lymphoma
- Hematopoietic Cell Transplantation for Non-Hodgkin Lymphomas
- Hematopoietic Cell Transplantation for Plasma Cell Dyscrasias, Including Multiple Myeloma and POEMS Syndrome

Benefit Application

Benefit determinations should be based in all cases on the applicable contract language. To the extent there are any conflicts between these guidelines and the contract language, the contract language will control. Please refer to the member's contract benefits in effect at the time of service to determine coverage or non-coverage of these services as it applies to an individual member.

Some state or federal mandates (e.g., Federal Employee Program [FEP]) prohibits plans from denying Food and Drug Administration (FDA)-approved technologies as investigational. In these instances, plans may have to consider the coverage eligibility of FDA-approved technologies on the basis of medical necessity alone.

Regulatory Status

The clonoSEQ® Minimal Residual Disease Test is offered by Adaptive Biotechnologies. ClonoSEQ® was previously marketed as ClonoSIGHT™ (Sequenta), which was acquired by Adaptive Biotechnologies in 2015. ClonoSIGHT™ was a commercialized version of the LymphoSIGHT platform by Sequenta for clinical use in MRD detection in lymphoid cancers. In September 2018, clonoSEQ received marketing clearance from the Food and Drug Administration through the de novo classification process to detect MRD in patients with Acute Lymphoblastic Leukemia (ALL) or Multiple myeloma (MM).

Rationale

Background Disease

There are 3 main types of hematologic malignancies: lymphomas, leukemias, and myelomas. Lymphoma is the most common type of hematologic malignancy and is typically divided into 2 categories, Hodgkin lymphoma (also known as Hodgkin disease) and non-Hodgkin lymphoma (NHL). Lymphoma begins in lymph cells of the immune system, which originate in bone marrow and collect in lymph nodes and other tissues. The 2 types of lymph cells that develop into NHL are B lymphocytes (B cells), which mature in the bone marrow, and T lymphocytes (T cells), which mature in the thymus.

Leukemia is caused by the overproduction of abnormal white blood cells in the bone marrow, which leads to a decrease in production of red blood cells and plasma cells. Leukemia may be acute or chronic, and affect either lymph or myeloid cells. The most common forms of leukemia are acute lymphoblastic leukemia (ALL), chronic lymphocytic leukemia (CLL), acute myeloid leukemia (AML), and chronic myeloid leukemia. There are a number of less common forms of leukemia. Multiple myeloma (MM), also called plasma myeloma, is a malignancy of plasma cells in the bone marrow.
Hodgkin Lymphoma
Hodgkin lymphoma is a relatively uncommon B-cell lymphoma. In 2017, the estimated number of new cases in the United States was approximately 8260 with 1070 estimated deaths. The disease has a bimodal distribution, with most patients diagnosed between the ages of 15 and 30 years, with a second peak in adults aged 55 years and older.

Non-Hodgkin Lymphoma
NHL includes a heterogeneous group of lymphoproliferative malignancies. In general, NHL can be divided into 2 prognostic groups: indolent and aggressive. Follicular lymphoma is the most common indolent NHL (70%-80% of cases), and often the terms indolent lymphoma and follicular lymphoma are used synonymously. Indolent NHL has a relatively good prognosis, with a median survival of 10 years; however, it is not curable in advanced clinical stages. Histologic transformation to higher grade lymphoma occurs in up to 70% of patients with low-grade lymphoma, and median survival with conventional chemotherapy is 1 year or less. Aggressive NHL has a shorter natural history; however, 30% to 60% of these patients can be cured with intensive combination chemotherapy regimens.

Acute Lymphoblastic Leukemia
Childhood ALL
ALL is the most common cancer diagnosed in children; it represents nearly 25% of cancers in children younger than 15 years. Remission of disease is now typically achieved with pediatric chemotherapy regimens in 98% of children with ALL, with up to 85% long-term survival rates. The prognosis after the first relapse is related to the length of the original remission. For example, the leukemia-free survival rate is 40% to 50% for children whose first remission was longer than 3 years compared with 10% to 15% for those who relapse less than 3 years after treatment.

Adult ALL
ALL accounts for 20% of acute leukemias in adults. Between 60% and 80% of adults with ALL can be expected to achieve a complete response after induction chemotherapy; however, only 35% to 40% can be expected to survive 2 years. “Poor prognosis” genetic abnormalities such as the Philadelphia chromosome (translocation of chromosomes 9 and 22) are seen in 25% to 30% of adult ALL but infrequently in childhood ALL. Other adverse prognostic factors in adult ALL include age greater than 35 years, poor performance status, male sex, and leukocytosis count of greater than 30,000/µL (B-cell lineage) or greater than 100,000/µL (T-cell lineage) at presentation.

Chronic Lymphocytic Leukemia
CLL tends to present as asymptomatic enlargement of the lymph nodes and tends to be indolent, but can undergo transformation to a more aggressive form of the disease. The median age at diagnosis of CLL is approximately 72 years. Both low- and intermediate-risk CLL demonstrate relatively good prognoses, with a median survival of 6 to 10 years; however, the median survival of high-risk CLL may only be 2 years. Although typically responsive to initial therapy, CLL is rarely cured by conventional therapy, and nearly all patients die of their disease.

Acute Myeloid Leukemia
AML, also called acute nonlymphocytic leukemia, refers to a set of leukemias that arise from a myeloid precursor in the bone marrow. Clinical signs and symptoms are associated with neutropenia, thrombocytopenia, and anemia. The incidence of AML increases with age, with a median of 67 years. Molecular studies have identified a number of genetic abnormalities that can be used to guide prognosis and management of AML. Cytogenetically normal AML is the largest defined subgroup of AML, comprising approximately 45% of all AML cases. Despite the absence of cytogenetic abnormalities, these cases often have genetic variants that affect outcomes.

Chronic Myeloid Leukemia
Chronic myeloid leukemia accounts for about 15% of newly diagnosed cases of leukemia in adults and occurs in 1 to 2 cases per 100,000 adults. The natural history of the disease consists of
an initial (indolent) chronic phase, lasting a median of 3 years, which typically transforms into an accelerated phase, followed by a “blast crisis,” which is usually the terminal event. Most patients present in chronic phase, often with nonspecific symptoms secondary to anemia and splenomegaly. Conventional-dose chemotherapy regimens used for chronic phase disease can induce multiple remissions and delay the onset of blast crisis to a median of 4 to 6 years. However, successive remissions are invariably shorter and more difficult to achieve than their predecessors.

**Multiple Myeloma**

MM represents approximately 10% of all hematologic cancers. It is treatable but rarely curable. Treatment is usually reserved for patients with symptomatic disease (usually progressive myeloma), whereas asymptomatic patients are observed because there is little evidence that early treatment of asymptomatic MM prolongs survival compared with therapy delivered at the time of symptoms or end-organ damage. In some patients, an intermediate asymptomatic but the more advanced premalignant stage is recognized and referred to as smoldering MM. The overall risk of disease progression from smoldering to symptomatic MM is 10% per year for the first 5 years, approximately 3% per year for the next 5 years, and 1% for the next 10 years.

**Treatment**

Treatment depends on the type of malignancy and may include surgery, radiotherapy, chemotherapy, targeted therapy, plasmapheresis, biologic therapy, or hematopoietic cell transplant. Treatment of the acute leukemias can lead to complete remission. MM and the chronic leukemias are treatable but generally incurable. Patients are typically followed by complete blood count and morphologic assessment of bone marrow. Complete hematologic response is defined as a bone marrow blast (immature cells) composition of less than 5% and hematologic recovery (normal neutrophil and platelet count) without the need for red blood cell transfusions.

**Measurable Residual Disease**

Relapse is believed to be due to residual clonal cells that remain following “complete response” after induction therapy but are below the limits of detection using conventional morphologic assessment. Residual clonal cells that can be detected in blood or bone marrow are referred to as measurable residual disease (MRD), also known as minimal residual disease. MRD assessment is typically performed by flow cytometry or polymerase chain reaction (PCR) with primers for common variants. Flow cytometry evaluates blasts based on the expression of characteristic antigens, while PCR assesses specific chimeric fusion gene transcripts, gene variants, and overexpressed genes. PCR is sensitive for specific targets, but clonal evolution may occur between diagnosis, treatment, remission, and relapse that can affect the detection of MRD.

Next-generation sequencing (NGS) has 10-to-100-fold greater sensitivity for detecting clonal cells (see Table 1) and does not require patient-specific primers. For both PCR and NGS a baseline sample at the time of high disease load is needed to identify tumor-specific sequences. MRD with NGS is frequently used as a surrogate measure of treatment efficacy in drug development and is transitioning from “bench-to-bedside” for clinical use.

It is proposed that by using a highly sensitive and sequential MRD surveillance strategy, one could expect better outcomes when therapy is guided by molecular relapse rather than hematologic relapse. However, some patients may have hematologic relapse despite no MRD, while others do not relapse despite residual mutation-bearing cells. Age-related clonal hematopoiesis, characterized by somatic variants in leukemia-associated genes with no associated hematologic disease, further complicates the assessment of MRD. There is currently no consensus on which method provides clinically meaningful assessment of MRD. A 2018 international consensus paper recommended that flow cytometry presents a high enough sensitivity to be used in routine clinical practice, but for a more sensitive result and if MRD eradication is the goal for the selected patient, then allele-specific PCR should be used.1 It is notable that next-generation flow techniques have reached a detection limit of one in 10^5 cells, which is equal to PCR and approaches the limit of detection of NGS (see Table 1).
available test (clonoSEQ) uses both PCR and NGS to detect clonal DNA in blood and bone marrow. ClonoSEQ Clonality (ID) PCR assessment is performed when there is a high disease load (e.g., initial diagnosis or relapse) to identify dominant or “trackable” B- or T-cell sequences associated with the malignant clone. NGS is then used to monitor the presence and level of the associated sequences in follow-up samples. As shown in Table 1, NGS can detect clonal cells with greater sensitivity than either flow cytometry or PCR. It is not known whether the increase in sensitivity from 10^{-5} to 10^{-6} represents a clinically meaningful difference in MRD.

<table>
<thead>
<tr>
<th>Technique</th>
<th>Sensitivity</th>
<th>Blast per 100,000 Nucleated Cells</th>
</tr>
</thead>
<tbody>
<tr>
<td>Microscopy (complete response)</td>
<td></td>
<td>50,000</td>
</tr>
<tr>
<td>Multiparameter flow cytometry</td>
<td>10^{-4}</td>
<td>10</td>
</tr>
<tr>
<td>Next-generation flow cytometry</td>
<td>10^{-5}</td>
<td>1.0</td>
</tr>
<tr>
<td>Polymerase chain reaction</td>
<td>10^{-5}</td>
<td>1.0</td>
</tr>
<tr>
<td>Quantitative next-generation sequencing</td>
<td>10^{-5}</td>
<td>1.0</td>
</tr>
<tr>
<td>Next-generation sequencing</td>
<td>10^{-6}</td>
<td>0.1</td>
</tr>
</tbody>
</table>

**Literature Review**
Evidence reviews assess whether a medical test is clinically useful. A useful test provides information to make a clinical management decision that improves the net health outcome. That is, the balance of benefits and harms is better when the test is used to manage the condition than when another test or no test is used to manage the condition.

The first step in assessing a medical test is to formulate the clinical context and purpose of the test. The test must be technically reliable, clinically valid, and clinically useful for that purpose. Evidence reviews assess the evidence on whether a test is clinically valid and clinically useful. Technical reliability is outside the scope of these reviews, and credible information on technical reliability is available from other sources.

**Next-Generation Sequencing to Detect Measurable Residual Disease in B-Cell Acute Lymphoblastic Leukemia**

**Clinical Context and Test Purpose**
ALL is the most common cancer diagnosed in children; it represents nearly 25% of cancers in children younger than 15 years and 20% of acute leukemias in adults. Remission of disease is now typically achieved with pediatric chemotherapy regimens in 98% of children with ALL, with up to 85% long-term survival rates. The prognosis after the first relapse is related to the length of the original remission. For example, the leukemia-free survival rate is 40% to 50% for children whose first remission was longer than 3 years compared with 10% to 15% for those who relapse less than 3 years after treatment. Between 60% and 80% of adults with ALL can be expected to achieve a complete response (CR) after induction chemotherapy; however, only 35% to 40% can be expected to survive 2 years. “Poor prognosis” genetic abnormalities such as the Philadelphia chromosome (translocation of chromosomes 9 and 22) are seen in 25% to 30% of adult ALL but infrequently in childhood ALL. Other adverse prognostic factors in adult ALL include age greater than 35 years, poor performance status, male sex, and leukocytosis count of greater than 30,000/μL (B-cell lineage) or greater than 100,000/μL (T-cell lineage) at presentation.

Induction therapy aims to reduce the leukemic cell population below the cytological detection limit (about 10^{10} cells or 1 malignant cell for every 20 to 100 normal cells), but it is believed that remaining leukemic cells that are below the level of clinical and conventional morphologic detection lead to relapse if no further treatment were given.1,2 Consolidation and intensification therapy is intended to eradicate this residual disease. The type of post-remission therapy (chemotherapy or autologous or allogeneic hematopoietic cell transplantation [HCT]) depends on the expected rate of relapse and patient characteristics such as age and comorbidities. Bone marrow is examined every three to six months for a minimum of two years to determine clinical relapse. If a patient is in CR for seven to eight years they are considered cured. Most
children and up to one-half of adults will have prolonged disease-free survival, but up to 20 percent of adults will have a resistant disease, and a majority of adults and some children will eventually relapse and die of leukemia.3,4.

Measurable, or minimal residual disease (MRD) is used to assess the subclinical residual disease. Patients with detectable MRD have an increased risk of relapse, but the absolute risk varies depending on the timing of MRD evaluation, the sensitivity of the method used, and baseline characteristics of the patient and tumor.3,4 In addition, not all patients with MRD positivity will relapse clinically because some cells with abnormal markers may lack the ability to create disease. Other patients will relapse despite no detectable disease as a result of malignant progenitor cells that lack the initially identified markers. MRD is most commonly measured with polymerase chain reaction (PCR) and flow cytometry (FC).

MRD assays are routinely used in the clinical care of children and increasingly in adults with ALL, although the choice of tests may depend on how the results will impact patient care.3,4 FC may be preferred if there are plans to escalate care because results are rapidly available and the likelihood of relapse with this less sensitive test is high. PCR may be preferred to identify patients with a low risk of relapse when a reduction in treatment intensity is being considered. Some clinicians use more than one technique to minimize false-negative results, or at multiple time points to assess disease trajectory, and ongoing trials are evaluating whether children who demonstrate a rapid clearance of tumor cells during induction therapy may be candidates for less intensive therapy. In adults who have a high rate of relapse, MRD is being studied to identify patients who require intensified treatment. One drug (blinatumomab) has received approval from the U.S. Food and Drug Administration to treat MRD positive B-cell precursor ALL with MRD positivity of 0.1% or greater (10⁻³).5

Next-generation sequencing (NGS) is a newer technique that is commercially available (e.g., ClonoSEQ). NGS is more sensitive than other methods and can detect up to 1 leukemic cell in 1000000 cells if there is sufficient DNA in the sample (see Table 1), but other performance characteristics are not well established.

**Test Purpose**

The main use of measurement of MRD with NGS is to risk-stratify and inform treatment management.

Measures of MRD can be used to assess whether a patient has failed to fully respond to treatment or is progressing after responding to treatment. If a patient meets the criteria for nonresponse or for relapse, the clinical decision generally would be to provide additional therapy prior to transplant. The analytic framework for the use of MRD for ALL, based on guidelines from the National Comprehensive Cancer Network, is shown in Figure 1.
The question addressed in this evidence review is: Does the use of NGS for MRD at different thresholds (e.g., $10^{-4}$ or $<10^{-4}$) improve the net health outcome in patients with B cell-ALL (B-ALL)? The following PICOs were used to select literature to inform this review.

**Patients**
The relevant population of interest are patients who have received induction therapy for B-ALL (see Figure 1). Patients who achieve a clinical CR following induction therapy would be assessed for MRD to determine whether additional therapy might be recommended prior to HCT. Patients who have relapsed or refractory diseases would be assessed for the Philadelphia chromosome and if negative may undergo assessment for MRD.

**Interventions**
The test being considered is MRD assessment by NGS (e.g., ClonoSEQ). This test is proposed as an adjunct to clinical assessment and an alternative to FC and PCR. NGS utilizes locus-specific primers for immunoglobulin gene rearrangements in IGH-VDJH, IGHDJH, or IGK. This technique does not require the use of patient-specific primers, but baseline bone marrow samples are required in order to identify the dominant clonotype. MRD positivity or negativity is reported at all thresholds (e.g., positive at $10^{-4}$ but negative at $10^{-5}$). The sensitivity of this technique can reach up to $10^{-6}$ depending on the quantity of DNA available from the bone marrow sample. This evidence review will evaluate outcomes for NGS at different thresholds.

**Comparators**
The following tests are currently being used to inform treatment decisions for those with B-ALL: FC (sensitivity of $10^{-4}$) and PCR (sensitivity of $10^{-5}$). Meta-analysis of 39 studies (13637 patients) that evaluated survival outcomes found that MRD negativity with either FC or PCR was associated with a better long-term outcome. Ten-year event-free survival with MRD negativity was 77% in children and 64% in adults compared to 32% and 21%, respectively, in patients who were MRD positive. For reference, the event-free survival hazard ratio (HR) for MRD negativity/positivity with FC or PCR was 0.23 (95% Bayesian credible interval 0.18-0.28) for pediatric patients and 0.28 (95% Bayesian credible interval, 0.24-0.33) for adults.

**Outcomes**
The general outcomes of interest are remission and relapse in the short-term and survival at a longer follow-up.
Beneficial outcomes of a true-positive test result (presence of clinically significant residual disease) would be the administration of an effective treatment leading to a reduction in relapse and improvement in overall survival (OS). The beneficial outcome of a true-negative test (absence of clinically significant disease) is the avoidance of unnecessary treatment and reduction of adverse events.

Harmful outcomes of a false-positive test are an unnecessary treatment for ALL resulting in treatment-related harms. Harmful outcomes of a false-negative test are a reduction in necessary treatment that would delay treatment, with a potential impact in progression-free survival (PFS) and OS.

Direct harms of the test are repeated bone marrow biopsy, although bone marrow samples are also needed for FC. Harms of repeated bone marrow biopsy may include tenderness or pain, bleeding or bruising, and swelling.

Relapse of B-ALL may be measured in two years. Changes in survival from B-ALL would be observable at a minimum of five years.

**Study Selection Criteria**
For the evaluation of the clinical validity of the ClonoSEQ test, studies that met the following eligibility criteria were considered:

- Included a suitable reference standard (PFS or OS)
- Evaluated outcomes at different levels of MRD or compared NGS to FC

OR, comparative trials that evaluated health outcomes when therapy was guided by NGS assessment of MRD

**Technically Reliable**
Assessment of technical reliability focuses on specific tests and operators and requires a review of unpublished and often proprietary information. Review of specific tests, operators, and unpublished data are outside the scope of this evidence review and alternative sources exist. This evidence review focuses on clinical validity and clinical utility.

**Clinically Valid**
A test must detect the presence or absence of a condition, the risk of developing a condition in the future, or treatment response (beneficial or adverse).

Tables 2, 3, and 4 describe studies that have evaluated prognosis based on MRD levels detected by FC and NGS. Overall, higher levels of MRD are associated with a worse prognosis. In a study by Wood et al (2018), there was high concordance between FC and NGS at a threshold of 10^{-4} in pediatric B-ALL (data are shown graphically in the publication). A subset of these results was submitted to the Food and Drug Administration in support of their de novo clearance. OS in pediatric patients with MRD positivity was significantly lower than in pediatric patients who were MRD negative at this threshold. At an MRD threshold of 10^{-4}, NGS identified 55 patients as MRD-positive who were MRD-negative by FC, while 17 patients were MRD-positive by FC but MRD-negative by NGS (see Table 3). Patients who were FC negative/NGS positive had outcomes that were midway between patients who were concordant as MRD positive or MRD negative for both tests.

Notably, higher levels of sensitivity were associated with a decrease in clinical specificity, with a larger fraction of MRD-positive patients with relatively good outcome (data not shown in the publication). With MRD negativity set at a threshold of 10^{-6}, OS was 100% in the standard-risk group and 95.1% in the high-risk group (see Table 4), but at this threshold, there was not a statistically significant difference in OS between the MRD positive and MRD negative patients for either group. The maximal HR for NGS was obtained at 10^{-4}, which is the sensitivity of FC. A smaller study by Pulsipher et al (2015) compared NGS at 10^{-6} with FC assessed before and after...
HCT in pediatric patients with ALL. NGS was more successful at predicting the relapse probability and OS compared to FC. The major limitations of these studies are shown in Tables 5 and 6. A limitation in Wood et al (2018) is that samples were only available at the end of induction, so the results only apply to the end of induction. In addition, the data on sensitivity and specificity at other thresholds were not reported, although the study did assess the threshold with the greatest HR, which was calculated to be $10^{-4}$ (the same as FC). Both studies were conducted in pediatric ALL patients, and results may not apply fully to adults or be applicable to other periods in the treatment course.

No studies were identified that evaluated the use of NGS to detect MRD in adult ALL.

Table 2. Characteristics of Prognostic Studies Assessing NGS for MRD

<table>
<thead>
<tr>
<th>Study</th>
<th>Study Population</th>
<th>Design</th>
<th>Reference Standard</th>
<th>Threshold for PIT</th>
<th>FU, y</th>
<th>Test Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wood et al (2018)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>619 paired bone marrow samples from pediatric B-ALL patients before and after induction chemotherapy in COG trials</td>
<td>Retrospective from banked samples with comparison of FC and NGS</td>
<td>Event-free survival and overall survival</td>
<td>FC at $10^{-4}$ NGS at $10^{-4}$ and $10^{-5}$</td>
<td>5</td>
<td>ImmunoSEQ</td>
</tr>
<tr>
<td>Pulsipher et al (2015)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Before (n=41) and after HCT (n=57) marrow samples from pediatric ALL patients in COG trials</td>
<td>Retrospective from banked samples with comparison of FC and NGS</td>
<td>Time to relapse following HCT</td>
<td>FC at $10^{-4}$ NGS at $10^{-6}$</td>
<td>5</td>
<td>ImmunoSEQ</td>
</tr>
</tbody>
</table>

ALL: acute lymphoblastic leukemia; COG: Children's Oncology Group; FC: flow cytometry; FU: follow-up; HCT: hematopoietic cell transplantation; MRD: measurable residual disease; NGS: next-generation sequencing; PIT: positive index test.

Table 3. Concordance Between FC and NGS at a Threshold of $10^{-4}$ from Wood et al (2018)<sup>a</sup>

<table>
<thead>
<tr>
<th>Flow Cytometry</th>
<th>+</th>
<th>-</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>NGS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+</td>
<td>87</td>
<td>55</td>
<td>142</td>
</tr>
<tr>
<td>-</td>
<td>17</td>
<td>409</td>
<td>426</td>
</tr>
<tr>
<td>Total</td>
<td>104</td>
<td>464</td>
<td>568</td>
</tr>
</tbody>
</table>

FC: flow cytometry; NGS: next-generation sequencing.

Table 4. Results of Prognostic Studies Assessing NGS for MRD

<table>
<thead>
<tr>
<th>Study</th>
<th>N</th>
<th>MRD Threshold</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Relapse Probability at 2 Years, %</td>
</tr>
<tr>
<td>Wood et al (2018)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>282</td>
<td>EOI NGS negative &lt;10^{-6} (n=56)</td>
<td>98.1 (2)</td>
</tr>
<tr>
<td></td>
<td>297</td>
<td>High Risk</td>
<td>EOI NGS negative &lt;10^{-6} (n=89)</td>
</tr>
<tr>
<td>Pulsipher et al (2015)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>41</td>
<td>Pre-HCTFC negative</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pre-HCTNGS negative</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pre-HCTFC positive</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pre-HCTNGS positive</td>
<td>53</td>
</tr>
</tbody>
</table>
Limitations in relevance and design and conduct are shown in Tables 4 and 5.

Table 5. Relevance Limitations

<table>
<thead>
<tr>
<th>Study</th>
<th>Populationa</th>
<th>Interventionb</th>
<th>Comparator</th>
<th>Outcomesd</th>
<th>Duration of FUe</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>are specific</td>
<td>ImmunoSEQ</td>
<td>ImmunoSEQ</td>
<td>FU</td>
<td></td>
</tr>
<tr>
<td></td>
<td>to pediatric</td>
<td>rather</td>
<td>rather</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>B-ALL</td>
<td>than ClonoSEQ</td>
<td>than ClonoSEQ</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Stored</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>samples</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>were</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>available</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>only at the</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>end of induction</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Pulsipher et al (2015)^b

4. Results are specific to pediatric ALL
3. Used ImmunoSEQ rather than ClonoSEQ

The study limitations stated in this table are those notable in the current review; this is not a comprehensive limitations assessment.

ALL: acute lymphoblastic leukemia; FU: follow-up.

Population key: 1. Intended use population unclear; 2. Clinical context is unclear; 3. Study population is unclear; 4. Study population not representative of intended use.

Intervention key: 1. Classification thresholds not defined; 2. Version used unclear; 3. Not intervention of interest.

Comparator key: 1. Classification thresholds not defined; 2. Not compared to credible reference standard; 3. Not compared to other tests in use for same purpose.

Outcomes key: 1. Study does not directly assess a key health outcome; 2. Evidence chain or decision model not explicated; 3. Key clinical validity outcomes not reported (sensitivity, specificity, and predictive values); 4. Reclassification of diagnostic or risk categories not reported; 5. Adverse events of the test not described (excluding minor discomforts and inconvenience of venipuncture or noninvasive tests).

Follow-Up key: 1. Follow-up duration not sufficient with respect to natural history of disease (true-positives, true-negatives, false-positives, false-negatives cannot be determined).

Table 6. Study Design and Conduct Limitations

<table>
<thead>
<tr>
<th>Study</th>
<th>Selectiona</th>
<th>Blindingb</th>
<th>Delivery of Testc</th>
<th>Selective Reportingd</th>
<th>Data Completenessf</th>
<th>Statisticalfl</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wood et al (2018)b</td>
<td>2. Selection based on availability of tissue samples from prior studies</td>
<td>2. NGS at 10^4 was not prespecified. The lack of specificity with other thresholds was mentioned</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pulsipher et al (2015)^b</td>
<td>2. Selection based on availability of tissue samples from prior studies</td>
<td>1. Blinding was not described</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

NGS: next-generation sequencing.

The study limitations stated in this table are those notable in the current review; this is not a comprehensive limitations assessment.
Section Summary: Clinical Validity
Evidence on the clinical validity of NGS to risk-stratify patients includes two retrospective studies in pediatric patients with ALL who had participated in earlier trials by the Children's Oncology Group. The largest study was conducted in stored samples from before and after induction therapy, and MRD negativity was one of several factors that were used to risk-stratify patients. Comparison with FC showed comparable results when the same threshold ($10^{-4}$) was used for both NGS and FC, and OS in pediatric patients with MRD positivity was significantly lower than in pediatric patients who were MRD negative. However, NGS at the limit of detection ($10^{-6}$ or 1 leukemic cell in 1000000 normal cells) was found to have lower specificity. Thus, in one study of over 600 pediatric patients with B-ALL undergoing induction, risk stratification based on NGS and FC were comparable at a threshold of $10^{-4}$, but NGS had more false-positives with lower thresholds.

Clinically Useful
A test is clinically useful if the use of the results informs management decisions that improve the net health outcome of care. The net health outcome can be improved if patients receive correct therapy, or more effective therapy, or avoid unnecessary therapy, or avoid unnecessary testing.

Direct Evidence
Direct evidence of clinical utility is provided by studies that have compared health outcomes for patients managed with and without the test. Because these are intervention studies, the preferred evidence would be from randomized controlled trials (RCTs). No trials were identified that compared outcomes when treatment was guided by NGS.

Chain of Evidence
Indirect evidence on clinical utility rests on clinical validity. If the evidence is insufficient to demonstrate test performance, no inferences can be made about clinical utility.

There is sufficient evidence on test performance when results of the NGS are reported at $10^{-4}$, which is comparable to other established methods of measuring MRD such as FC. However, performance characteristics at lower thresholds are uncertain, and there is some evidence that false-positives may be increased with a more sensitive test. Therefore, a chain of evidence cannot be constructed regarding the clinical utility of measurement of MRD at less than $10^{-4}$ in patients with ALL.

Section Summary: NGS to Detect MRD in ALL
Evidence is sufficient to support the clinical utility of using NGS to measure MRD when patient management is based on test results at a sensitivity of $10^{-4}$. Evidence is insufficient to evaluate benefits and harms when treatment decisions are made based on NGS results at thresholds lower than $10^{-4}$. Few studies have been performed to assess whether the identification of 1 in 1000000 cells identifies clinically significant residual disease, and false-positives may be increased resulting in harm from overtreatment. Further study is needed to clarify which threshold of NGS should be considered when risk stratifying patients and whether treatment decisions based on the more sensitive assay improves the net health outcome.
Next-Generation Sequencing to Detect Measurable Residual Disease in Multiple Myeloma

Clinical Context and Test Purpose

MM represents approximately 17% of all hematologic cancers, largely occurring in patients over 60. It is characterized by the proliferation of plasma cells in the bone marrow producing a monoclonal immunoglobulin. The clonal plasma cells frequently result in extensive skeletal destruction with osteolytic lesions, osteopenia, and/or pathologic fractures; additional complications can include hypercalcemia, renal insufficiency, anemia, and infections.10

MM is treatable but is typically incurable, with treatment reserved for patients with the symptomatic disease (usually progressive). Without effective therapy, symptomatic patients die within a median of six months. Asymptomatic patients are observed because there is little evidence that early treatment of asymptomatic MM prolongs survival compared with therapy delivered at the time of symptoms or end-organ damage. In some patients, an asymptomatic but more advanced premalignant stage is referred to as smoldering MM. Patients with smoldering MM may remain stable for prolonged periods, with an overall risk of disease progression from smoldering to symptomatic MM of 10% per year for the first five years, approximately 3% per year for the next five years, and 1% for the next ten years.

Prognosis and treatment for MM depend on risk stratification based on underlying genetic variants, age, performance status, comorbidities, stage, and response to therapy. Patients are assessed to determine eligibility for HCT because HCT has been shown to prolong both event-free and OS compared with chemotherapy alone. The response to treatment is usually determined by a morphologic evaluation and visual quantitation of the percentage of plasma cells in the bone marrow. Most patients with MM will have an initial response to treatment, but will ultimately progress with serial relapse, and will be treated with most available agents at some point during their disease course. Other patients will not respond to initial treatment (refractory disease).

Response to treatment is categorized into clinical response, MRD response, and imaging response. A complete (clinical) response is defined by the International Myeloma Working Group and the National Comprehensive Cancer Network as shown in Table 7.11,12 MRD response is defined as a CR plus the absence of clonal plasma cells by next-generation flow (NGF) or NGS at a minimum sensitivity of 1 in 10^5 nucleated cells in bone marrow, and there is a category of “imaging plus MRD-negative” in which patients are determined to have a CR, be MRD negative in the bone marrow, and have also achieved PET/CT-negativity. “Sustained MRD negativity” is achieved when both imaging plus MRD are negative in assessments that are a minimum of one year apart. It is not known whether patients with sustained MRD negative status can be considered cured. MRD measured by NGS is currently used as a surrogate outcome measure in clinical trials, and there are ongoing trials to test the effectiveness of using NGS-MRD to guide therapy.13

<table>
<thead>
<tr>
<th>Table 7. Definitions of Complete Response and Measurable Residual Disease Criteria from the International Myeloma Working Group11.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Standard Response Criteria</strong></td>
</tr>
<tr>
<td>Complete response</td>
</tr>
<tr>
<td>MRD Response Criteria (requires a complete response)</td>
</tr>
<tr>
<td>Sequencing MRD-negative</td>
</tr>
<tr>
<td>Imaging plus MRD-negative</td>
</tr>
</tbody>
</table>

MRD: minimal residual disease; NGF: next-generation flow; NGS: next-generation sequencing
**Test Purpose**

The main use of measurement of MRD is to inform treatment management. Measures of MRD can be used to assess whether a patient has responded to treatment, has not fully responded to treatment, or has progressed. The analytic framework for the use of MRD for MM, based on guidelines from the National Comprehensive Cancer Network 12, is shown in Figure 2. If a patient meets the criteria for CR and MRD, the patient could proceed to maintenance therapy or observation. If, however, a patient meets the criteria for nonresponse or progression, the clinical decision would be to proceed to the next line of therapy for the previously treated disease. The National Comprehensive Cancer Network guidelines recommend guiding treatment based on multiparameter FC (threshold of 10^-4), with NGF or NGS used for prognosis at a threshold of 10^-5 or 10^-6. NGF is not widely performed in the U.S.

**Figure 2. Analytic Framework for the use of MRD to Inform Treatment Management in MM.**

Clinical pathways subsequent to treatment failure or relapse incorporate an accumulation of clinical trial evidence and are codified in clinical guidelines.

The question addressed in this evidence review is: Is the net health outcome improved when treatment is guided by MRD measured by NGS in patients with MM?

The following PICOs were used to select literature to inform this review.

**Patients**

The relevant population of interest are patients who are undergoing or have undergone treatment for MM.

**Interventions**

The test being considered is MRD assessment by NGS (e.g., ClonoSEQ). NGS utilizes locus-specific primers for immunoglobulin gene rearrangements, which are rearranged in myeloma patients. Baseline bone marrow samples at the time of high disease load are required in order to identify the dominant clonotype. With the ClonoSEQ test, dominant ("clonogenic") sequences can be identified in ~92% of MM patients, while dominant sequences cannot be identified in the other ~8% of patients.

**Comparators**

Evaluation for disease progression in MM typically includes serum protein electrophoresis, serum immunofixation, 24-hour urine protein electrophoresis, urine immunofixation, and serum-free light chain, hemoglobin, serum calcium, and creatinine. A bone marrow aspirate and biopsy is not always needed but can clarify disease status and determine if a change in the cytogenetic characteristics has occurred. MRD detection by NGS would be an adjunct to clinical measures of progression and an alternative to FC, which has a sensitivity of 10^-4.
Outcomes
The general outcomes of interest are a clinical progression in the short term and survival at a longer follow-up.

Beneficial outcomes of a true-positive test result (detection of clinically significant disease) would be intensification or continuation of an effective treatment leading to longer PFS. The beneficial outcome of a true-negative test (absence of clinically significant residual disease) is the avoidance of unnecessary treatment and reduction of adverse events.

Harmful outcomes of a false-positive test include an increase or continuation of unnecessary treatment resulting in treatment-related harm. Harmful outcomes of a false-negative test include a reduction in necessary treatment that would delay treatment, with a potential impact in disease progression.

Direct harms of the test are repeated bone marrow biopsy. Harms of repeated bone marrow biopsy may include tenderness or pain, bleeding or bruising, and swelling.

Utility of MRD to guide treatment of MM may be measured in months for progression of the disease, with survival measured in years.

Study Selection Criteria
For the evaluation of the clinical validity of the ClonoSEQ test, studies that met the following eligibility criteria were considered:

- Included a suitable reference standard (PFS or OS)
- Evaluated outcomes at different levels of MRD

OR, comparative trials that evaluated health outcomes when therapy was guided by NGS assessment of MRD.

Technically Reliable
Assessment of technical reliability focuses on specific tests and operators and requires a review of unpublished and often proprietary information. Review of specific tests, operators, and unpublished data are outside the scope of this evidence review and alternative sources exist. This evidence review focuses on the clinical validity and clinical utility.

Clinical Validity
A test must detect the presence or absence of a condition, the risk of developing a condition in the future, or treatment response (beneficial or adverse).

Two published retrospective studies were identified that evaluated the association between MRD by NGS and disease progression in patients with MM (see Table 8). Both studies assessed MRD levels from patients who had participated in earlier MM treatment trials (the GEM myeloma trials and the IFM 2009 trial). Martínez-López et al (2014) assessed the time to progression (TTP) stratified by MRD at levels from $10^{-3}$ to $<10^{-5}$ and found that the TTP was associated with the level of MRD. Specifically, median progression was 27 months for patients with MRD $>10^{-3}$, 48 months for patients with MRD between $10^{-3}$ and $10^{-5}$, and 80 months for patients with MRD $<10^{-5}$, giving a HR of 3.97 for higher levels of MRD ($p<0.001$, see Table 9). In the subgroup of patients with CR, TTP was 131 months in MRD negative patients and 35 months in MRD positive patients (HR of 2.87, $p<0.001$).

When compared to multiparameter FC, 82 of 99 results (83%) were concordant (see Table 10). NGS identified an additional 12 patients with MRD that were MRD-negative by FC, while 5 patients were found to be flow MRD+/NGS MRD- (see Table 10). One of five flow+/NGS- patients progressed. Patients who were NGs+/flow- had an intermediate TTP (50 months) compared to NGs-negative patients (TTP not reached; $P<0.0001$).
In Perrot et al (2018), a threshold of $10^{-6}$ was used to evaluate the association between MRD and PFS, finding that the dichotomous division into MRD positive and MRD negative (no detectable MRD at the limit of detection) was highly predictive of PFS with an HR for MRD negative/MRD positive of 0.19 ($p<0.001$). The median PFS was 29 months in patients who were positive for MRD and was not reached among patients with no detectable MRD.

The major limitations of these studies are described in Tables 11 and 12. The report by Perrot et al (2018) was described as exploratory analysis, and in the study by Martinez-Lopez et al (2014), it does not appear that analysis by the level of MRD was pre-specified. In addition, the sample size in the study by Martinez-Lopez et al (2014) was limited by the availability of stored tissue from the earlier clinical studies. Perrot et al (2018) had a larger sample from the IFM 2009 trial but did not separately assess the subgroup of patients who had CR. Although Perrot et al (2018) also had a high number of cases that were not available for assessment, the publication supplement included a sensitivity analysis to examine the effect of missing data. The analysis was conducted using multiple imputations to impute missing MRD values. The results from the sensitivity analyses were reported to be similar to the primary results.

Similar results were obtained in the retrospective studies submitted for the Food and Drug Administration de novo application using a threshold of $10^{-5}$ to determine MRD negativity. In the 75 patients who had a CR in the phase 3 DFCI 10-106 study, continuous levels of MRD were marginally associated with PFS ($p=0.064$). Analysis from all participants in the phase 3 ALCYONE trial found that MRD negativity was associated with longer PFS (patients without a CR were considered MRD positive), but results were not reported separately for patients with CR.

### Table 8. Characteristics of Studies Assessing NGS for MRD

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Martinez-Lopez et al (2014)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Patients with available bone marrow samples from GEM myeloma trials&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Retrospective</td>
<td>TTP</td>
<td>MRD at $10^{-3}$ and $10^{-5}$</td>
<td>LymphoSIGHT</td>
</tr>
<tr>
<td>Perrot et al (2018)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Patients with myeloma enrolled in the IFM 2009 clinical trial&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Retrospective</td>
<td>PFS and OS</td>
<td>MRD at $10^{-6}$</td>
<td>ClonoSEQ</td>
</tr>
</tbody>
</table>

MRD: measurable residual disease; NGS: next-generation sequencing; OS: overall survival; PFS: progression-free survival; TTP: time to progression.

<sup>a</sup> GEM (Grupo Español de Mieloma) myeloma treatment trials

<sup>b</sup> IFM 2009 was phase 3 trial from the Intergroupe Francophone du Myelome, conducted between 2010 and 2012, which evaluated the role of autologous cell transplantation in patients with newly diagnosed myeloma.

### Table 9. Results of Prognostic Studies Assessing NGS for MRD

<table>
<thead>
<tr>
<th>Study</th>
<th>N</th>
<th>MRD Threshold</th>
<th>TTP, mo (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Martinez-Lopez et al (2014)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>133</td>
<td>$&gt;10^{-3}$</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10^{-3} to $10^{-5}$</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$&lt;10^{-5}$</td>
<td>80</td>
</tr>
</tbody>
</table>

Hazard Ratio for Time to Progression

<table>
<thead>
<tr>
<th>Study</th>
<th>Hazard Ratio for Time to Progression</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Martinez-Lopez et al (2014)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>3.97</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Subset of patients with CR

<table>
<thead>
<tr>
<th>Study</th>
<th>Hazard Ratio for Time to Progression</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subset of patients with CR</td>
<td>131 (51-154)</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Study</th>
<th>Hazard Ratio for Progression Free Survival (95% CI)</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perrot et al (2018)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.19 (0.13 to 0.26)</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

CI: confidence interval; CR: complete response; MRD: measurable residual disease; NGS: next-generation sequencing; TTP: time to progression.
Table 10. Concordance Between NGS and FC in Study by Martinez-Lopez (2014)

<table>
<thead>
<tr>
<th>Flow Cytometry</th>
<th></th>
<th></th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>NGS +</td>
<td>60</td>
<td>12</td>
<td>72</td>
</tr>
<tr>
<td>NGS -</td>
<td>5</td>
<td>22</td>
<td>27</td>
</tr>
<tr>
<td>Total</td>
<td>65</td>
<td>34</td>
<td>99</td>
</tr>
</tbody>
</table>

FC: flow cytometry; NGS: next-generation sequencing.

Table 11. Relevance Limitations

<table>
<thead>
<tr>
<th>Study</th>
<th>Populationa</th>
<th>Interventionb</th>
<th>Comparatorc</th>
<th>Outcomesd</th>
<th>Duration of FUe</th>
</tr>
</thead>
<tbody>
<tr>
<td>Martinez-Lopez et al (2014)14</td>
<td>3. No data were reported using a threshold of 10-6 since most of the samples had less input cells than is needed for this level of sensitivity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perrot et al (2018)15</td>
<td>4. The study included patients from the IFM 2009 trial who had at least a very good partial response but did not report separately on patients with a complete response</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The study limitations stated in this table are those notable in the current review; this is not a comprehensive limitations assessment.

FU: follow-up.
a Population key: 1. Intended use population unclear; 2. Clinical context is unclear; 3. Study population is unclear; 4. Study population not representative of intended use.
b Intervention key: 1. Classification thresholds not defined; 2. Version used unclear; 3. Not intervention of interest.
c Comparator key: 1. Classification thresholds not defined; 2. Not compared to credible reference standard; 3. Not compared to other tests in use for same purpose.
d Outcomes key: 1. Study does not directly assess a key health outcome; 2. Evidence chain or decision model not explicated; 3. Key clinical validity outcomes not reported (sensitivity, specificity, and predictive values); 4. Reclassification of diagnostic or risk categories not reported; 5. Adverse events of the test not described (excluding minor discomforts and inconvenience of venipuncture or noninvasive tests).
e Follow-Up key: 1. Follow-up duration not sufficient with respect to natural history of disease (true-positives, true-negatives, false-positives, false-negatives cannot be determined).

Table 12. Study Design and Conduct Limitations

<table>
<thead>
<tr>
<th>Study</th>
<th>Selectiona</th>
<th>Blindingb</th>
<th>Delivery of Testc</th>
<th>Selective Reportingd</th>
<th>Data Completenesse</th>
<th>Statisticalf</th>
</tr>
</thead>
<tbody>
<tr>
<td>Martinez-Lopez et al (2014)14</td>
<td>2. Selection based on availability of tissue samples from prior studies</td>
<td>1. Blinding not described</td>
<td></td>
<td></td>
<td>1. The analysis by level of MRD does not appear to be prespecified.</td>
<td></td>
</tr>
<tr>
<td>Perrot et al (2018)15</td>
<td>2. Selection based on availability of tissue samples in the original study</td>
<td>1. Blinding not described</td>
<td></td>
<td></td>
<td>1. Post-hoc exploratory analysis, not adjusted for multiple comparisons</td>
<td></td>
</tr>
</tbody>
</table>

MRD: measurable residual disease
The study limitations stated in this table are those notable in the current review; this is not a comprehensive limitations assessment.

a Selection key: 1. Selection not described; 2. Selection not random or consecutive (i.e., convenience).
b Blinding key: 1. Not blinded to results of reference or other comparator tests.
c Test Delivery key: 1. Timing of delivery of index or reference test not described; 2. Timing of index and comparator tests not same; 3. Procedure for interpreting tests not described; 4. Expertise of evaluators not described.
e Data Completeness key: 1. Inadequate description of indeterminate and missing samples; 2. High number of samples excluded; 3. High loss to follow-up or missing data.
f Statistical key: 1. Confidence intervals and/or p values not reported; 2. Comparison with other tests not reported.

Section Summary: Clinical Validity

The evidence on NGS for detection of MRD includes two published retrospective studies and additional retrospective studies from the Summary of Safety and Effectiveness of the de novo application for ClonoSEQ in patients with MM. These studies evaluated the association between the level of MRD detected by NGS in the bone marrow and the TTP from the completed phase 3 trials. All of the studies demonstrated an association between the level of MRD and PFS with longer TTP in patients who exhibit MRD negativity below 10^{-5} or 10^{-6} compared to patients who have detectable residual disease. There was also high concordance between NGS and FC. Patients who were discordant for the two tests had outcomes that were intermediate between patients who were positive for both tests and those who were negative for both tests.

Clinically Useful

A test is clinically useful if the use of the results informs management decisions that improve the net health outcome of care. The net health outcome can be improved if patients receive correct therapy, or more effective therapy, or avoid unnecessary therapy, or avoid unnecessary testing.

Direct Evidence

Direct evidence of clinical utility is provided by studies that have compared health outcomes for patients managed with and without the test. Because these are intervention studies, the preferred evidence would be from RCTs.

No RCTs assessing the clinical utility of MRD by NGS to guide therapy were identified.

Chain of Evidence

Indirect evidence on clinical utility rests on clinical validity. High concordance has been shown between NGS and FC at a threshold of 10^{-4}, indicating that NGS may be considered an alternative to FC at this threshold.

The retrospective studies are insufficient to demonstrate clinical validity at thresholds lower than 10^{-4}. Levels of MRD are associated with average prognosis, but performance characteristics are unknown at the level of sensitivity that is provided by NGS. A potential benefit of NGS assessment of MRD would be if patients were able to forgo maintenance therapy if there was no detectable MRD. However, it is unknown whether therapy can be safely eliminated based on this test.

Section Summary: Clinically Useful

In an exploratory analysis of the largest study to date, the median PFS was 29 months in patients who were positive for MRD and was not reached among patients with no detectable clones, suggesting that assessment of MRD might have utility in guiding therapy. Although there is high concordance between FC and NGS at a threshold of 10^{-4}, performance characteristics, such as the rate of false-positives, at more sensitive thresholds are unknown. About one-quarter of MRD negative patients progressed within 36 months in these trials, raising questions about whether NGS could be used to guide therapy. It is unknown whether progression is due to very low levels of residual disease or to new clonal rearrangements in MM. Direct evidence from RCTs is needed.
to evaluate whether patient outcomes are improved by changes in postinduction care (e.g., continuing or discontinuing therapy, avoiding unnecessary adverse events) following NGS assessment of residual disease. Several trials that will test the effectiveness of MRD to guide therapy in MM are ongoing.

**Summary of Evidence**

For individuals with B-ALL who are being monitored for residual disease following treatment who receive NGS for MRD at a threshold of $10^{-4}$, the evidence includes a retrospective comparison of data from two earlier trials by the Children's Oncology Group. The relevant outcomes are OS, disease-specific survival, test validity, change in disease status, quality of life (QOL), and treatment-related morbidity. Comparison of NGS and the established standard of FC showed good concordance when the same threshold ($10^{-4}$) was used for both NGS and FC. OS in pediatric patients with MRD positivity was significantly lower than in pediatric patients who were MRD negative at this threshold. The relatively small subset of patients who were discordant for FC and NGS results had outcomes that were midway between patients who were concordant as MRD positive or MRD negative for both tests. As the vast majority of patients had concordant results for NGS and FC at a threshold of $10^{-4}$, NGS can be considered an alternative to FC for monitoring MRD in patients with B-ALL. The evidence is sufficient to determine that the technology results in a meaningful improvement in the net health outcome.

For individuals with B-ALL who are being monitored for residual disease following treatment who receive NGS for MRD at a threshold of less than $10^{-4}$, the evidence includes retrospective analysis of prognosis from the earlier Children's Oncology Group trials. The relevant outcomes are OS, disease-specific survival, test validity, change in disease status, QOL, and treatment-related morbidity. NGS can be more sensitive than FC to detect the presence of residual leukemic cells, but specificity may be decreased at the more sensitive thresholds resulting in potential harm from overtreatment. Further study is needed to clarify whether MRD at levels lower than 1 in 10000 cells represents clinically significant disease and if the more sensitive test can be used to risk-stratify patients with B-ALL. The evidence is insufficient to determine the effects of the technology on health outcomes.

For individuals with MM who have achieved a CR following treatment who receive NGS for MRD at a threshold of $10^{-4}$, the evidence includes a retrospective comparison of NGS and FC data from MM treatment trials. The relevant outcomes are OS, disease-specific survival, test validity, change in disease status, QOL, and treatment-related morbidity. Comparison of NGS and the established standard of FC at $10^{-4}$ show good concordance. PFS in patients with MRD positivity is significantly shorter than in patients who are MRD negative at this threshold. The relatively small subset of patients who were discordant for FC and NGS results had outcomes that were, on average, midway between patients who were concordant as MRD positive or MRD negative for both tests. As the vast majority of patients had concordant results for NGS and FC at a threshold of $10^{-4}$, NGS can be considered an alternative to FC for monitoring MRD in patients with MM. The evidence is sufficient to determine that the technology results in a meaningful improvement in the net health outcome.

For individuals with MM who have achieved a CR following treatment who receive NGS for MRD at a threshold of less than $10^{-4}$, the evidence includes retrospective studies on prognosis. There is some evidence that MRD may be a prognostic marker, but there is insufficient evidence on the number of false-positives in patients with CR at the more sensitive threshold provided by NGS to guide therapy. A chain of evidence regarding management changes based on the assessment of MRD with NGS to detect 1 malignant clonal sequence out of 1000000 cells cannot be completed. Direct evidence from RCTs is needed to evaluate whether patient outcomes are improved by changes in postinduction care (e.g., continuing or discontinuing therapy, avoiding unnecessary adverse events) following NGS assessment of residual disease at a threshold lower than $10^{-4}$. Several trials that will test the
effectiveness of NGS to guide therapy in MM are ongoing. The evidence is insufficient to determine the effects of the technology on health outcomes.

**Supplemental Information**

**Practice Guidelines and Position Statements**

**International Myeloma Working Group**

The International Myeloma Working Group developed consensus criteria for response and minimal residual disease assessment in multiple myeloma (see Table 13).¹¹

**Table 13 IMWG Criteria**

<table>
<thead>
<tr>
<th>Standard Response Criteria</th>
<th>Complete response</th>
<th>Stringent complete response</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>“Negative immunofixation on the serum and urine and disappearance of any soft tissue plasmacytomas and &lt;5% plasma cells in bone marrow aspirates”</td>
<td>“Complete response as defined below plus normal FLC ratio** and absence of clonal cells in bone marrow biopsy by immunohistochemistry (κ/λ ratio ≤4:1 or ≥1:2 for κ and λ patients, respectively, after counting ≥100 plasma cells)”</td>
</tr>
</tbody>
</table>

**MRD Response Criteria (requires a complete response)**

<table>
<thead>
<tr>
<th>Sequencing MRD-negative</th>
<th>Absence of clonal plasma cells by NGS using the Lympho SIGHT platform (or validated equivalent) with a minimum sensitivity of 1 in 10^µ nucleated cells</th>
</tr>
</thead>
<tbody>
<tr>
<td>Imaging plus MRD-negative</td>
<td>MRD negativity by NGF or NGS plus imaging criteria</td>
</tr>
</tbody>
</table>

| Sustained MRD-negative | MRD negativity by NGF or NGS, and by imaging, at a minimum of 1 year apart. |

FLC: free light chain; IMWG: International Myeloma Working Group; MRD: minimal residual disease; NGF: next-generation flow; NGS: next-generation sequencing.

**The National Comprehensive Cancer Network**

The National Comprehensive Cancer Network has published guidelines of relevance to this review (see Table 14).

**Table 14. Recommendations on Assessing Measurable Residual Disease**

<table>
<thead>
<tr>
<th>Guideline</th>
<th>Version</th>
<th>Recommendation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acute lymphoblastic leukemia¹⁶</td>
<td>2.2019</td>
<td>Risk stratification after treatment induction by MRD positivity. MRD in ALL refers to the presence of leukemic cells below the threshold of detection by conventional morphologic methods. The most frequently employed methods for MRD assessment are FC, RQ-PCR, and NGS. The concordance rate between these methods is generally high.</td>
</tr>
<tr>
<td>Multiple myeloma¹²</td>
<td>2.2020</td>
<td>Bone marrow aspirate with multiparameter flow cytometry is to be used as clinically indicated following treatment. MRD tests should be initiated only at the time of suspected CR, and can be assessed for prognosis after a shared decision with the patient.</td>
</tr>
</tbody>
</table>

ALL: acute lymphoblastic leukemia, CR: complete response; FC: flow cytometry; MRD: measurable residual disease; NGS: next-generation sequencing; RQ-PCR: real-time quantitative polymerase chain reaction.

**U.S. Preventive Services Task Force Recommendations**

Not applicable.
Medicare National Coverage
Effective 01/17/2019, Molecular Diagnostic Services Program has determined that ClonoSEQ Assay testing is reasonable and necessary when performed on bone marrow specimens in patients with B-Cell ALL or multiple myeloma. Medicare will pay for a single episode of testing using ClonoSEQ for a patient with ALL or multiple myeloma when ClonoSEQ is being used according to its Food and Drug Administration cleared indications and clinical guidelines. An episode of testing will typically require a baseline assay and three follow-up assays.

Ongoing and Unpublished Clinical Trials
Some currently ongoing and unpublished trials that might influence this review are listed in Table 15.

Table 15. Summary of Key Trials

<table>
<thead>
<tr>
<th>NCT No.</th>
<th>Trial Name</th>
<th>Planned Enrollment</th>
<th>Completion Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ongoing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NCT03509961</td>
<td>A Phase II Pilot Trial to Estimate Survival After a Non-total Body Irradiation (TBI) Based Conditioning Regimen in Patients Diagnosed With Acute Lymphoblastic Leukemia (ALL) Who Are Pre-allogeneic Hematopoietic Cell Transplantation (HCT) Next-generation-sequence (NGS) Minimal Residual Disease (MRD) Negative (ENRAD)</td>
<td>95</td>
<td>Apr 2022</td>
</tr>
<tr>
<td>NCT03224507</td>
<td>Monoclonal Antibody-Based Sequential Therapy for Deep Remission in Multiple Myeloma - MASTER Trial</td>
<td>82</td>
<td>Apr 2023</td>
</tr>
<tr>
<td>NCT03914625</td>
<td>A Phase 3 Trial Investigating Blinatumomab (NSC # 765986) in Combination With Chemotherapy in Patients With Newly Diagnosed Standard Risk or Down Syndrome B-Lymphoblastic Leukemia (B-ALL) and the Treatment of Patients With Localized B-Lymphoblastic Lymphoma (B-LLy)</td>
<td>6720</td>
<td>Jun 2027</td>
</tr>
</tbody>
</table>

NCT: national clinical trial.

References

Documentation for Clinical Review

- No records required

Coding

This Policy relates only to the services or supplies described herein. Benefits may vary according to product design; therefore, contract language should be reviewed before applying the terms of the Policy. Inclusion or exclusion of codes does not constitute or imply member coverage or provider reimbursement.

MN/IE

The following services may be considered medically necessary in certain instances and investigational in others. Services may be considered medically necessary when policy criteria are met. Services may be considered investigational when the policy criteria are not met or when the code describes application of a product in the position statement that is investigational.

<table>
<thead>
<tr>
<th>Type</th>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPT®</td>
<td>81599</td>
<td>Unlisted multianalyte assay with algorithmic analysis</td>
</tr>
<tr>
<td></td>
<td>81479</td>
<td>Unlisted molecular pathology procedure</td>
</tr>
</tbody>
</table>

Policy History

This section provides a chronological history of the activities, updates and changes that have occurred with this Medical Policy.

<table>
<thead>
<tr>
<th>Effective Date</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>12/01/2018</td>
<td>BCBSA Medical Policy Adoption</td>
</tr>
<tr>
<td>12/01/2019</td>
<td>Policy revision without position change</td>
</tr>
<tr>
<td>02/02/2020</td>
<td>Policy statement, guidelines and literature updated</td>
</tr>
</tbody>
</table>

Definitions of Decision Determinations

Medically Necessary: Services that are Medically Necessary include only those which have been established as safe and effective, are furnished under generally accepted professional standards to treat illness, injury or medical condition, and which, as determined by Blue Shield, are: (a) consistent with Blue Shield medical policy; (b) consistent with the symptoms or diagnosis; (c) not furnished primarily for the convenience of the patient, the attending Physician or other provider; (d) furnished at the most appropriate level which can be provided safely and
effectively to the patient; and (e) not more costly than an alternative service or sequence of services at least as likely to produce equivalent therapeutic or diagnostic results as to the diagnosis or treatment of the Member’s illness, injury, or disease.

**Investigational/Experimental:** A treatment, procedure, or drug is investigational when it has not been recognized as safe and effective for use in treating the particular condition in accordance with generally accepted professional medical standards. This includes services where approval by the federal or state governmental is required prior to use, but has not yet been granted.

**Split Evaluation:** Blue Shield of California/Blue Shield of California Life & Health Insurance Company (Blue Shield) policy review can result in a split evaluation, where a treatment, procedure, or drug will be considered to be investigational for certain indications or conditions, but will be deemed safe and effective for other indications or conditions, and therefore potentially medically necessary in those instances.

### Prior Authorization Requirements (as applicable to your plan)

Within five days before the actual date of service, the provider must confirm with Blue Shield that the member's health plan coverage is still in effect. Blue Shield reserves the right to revoke an authorization prior to services being rendered based on cancellation of the member's eligibility. Final determination of benefits will be made after review of the claim for limitations or exclusions.

Questions regarding the applicability of this policy should be directed to the Prior Authorization Department at (800) 541-6652, or the Transplant Case Management Department at (800) 637-2066 ext. 3507708 or visit the provider portal at www.blueshieldca.com/provider.

Disclaimer: This medical policy is a guide in evaluating the medical necessity of a particular service or treatment. Blue Shield of California may consider published peer-reviewed scientific literature, national guidelines, and local standards of practice in developing its medical policy. Federal and state law, as well as contract language, including definitions and specific contract provisions/exclusions, take precedence over medical policy and must be considered first in determining covered services. Member contracts may differ in their benefits. Blue Shield reserves the right to review and update policies as appropriate.