ESMO consensus conference on malignant lymphoma: general perspectives and recommendations for prognostic tools in mature B-cell lymphomas and chronic lymphocytic leukaemia

M. Ladetto1*, C. Buske2, M. Hutchings3, M. Dreyling4, G. Gaidano5, S. Le Gouill6, S. Luminari7,8, C. Pott9, A. Zamlo10, E. Zucca11 & the ESMO Lymphoma Consensus Conference Panel Members†

1Hematology Division, Azienda Ospedaliera Santi Antonio e Biagio e Cesare Arrigo, Alessandria, Italy; 2Comprehensive Cancer Center Ulm and Department of Internal Medicine III, Institute of Experimental Cancer Research, University Hospital, Ulm, Germany; 3Department of Hematology, Rigshospitalet, Copenhagen, Denmark; 4Medizinische Klinik III, Klinikum der Universität München/LMU, Munich, Germany; 5Division of Hematology, Department of Translational Medicine, Amedeo Avogadro University of Eastern Piedmont, Novara, Italy; 6Clinical Hematology, Centre Hospitalo-Universitaire de Nantes, UMR892 Team 10, CIC Nantes, France; 7Hematology, Arcispedale S. Maria Nuova, IRCCS Reggio Emilia; 8Department of Diagnostic, Clinical and Public Health Medicine, University of Modena and Reggio Emilia, Modena, Italy; 9Second Medical Department, University Hospital Schleswig-Holstein, Kiel, Germany; 10Department of Diagnostics and Public Health, University of Verona, Verona, Italy; 11Lymphoma Unit, Oncology Institute of Southern Switzerland, Ospedale San Giovanni, Bellinzona, Switzerland

Received 9 June 2016; revised 11 August 2016; accepted 23 August 2016

The European Society for Medical Oncology (ESMO) consensus conference on mature B-cell lymphomas and chronic lymphocytic leukaemia (CLL) was held on 20 June 2015 in Lugano, Switzerland, and included a multidisciplinary panel of 25 leading experts. The aim of the conference was to develop recommendations on critical subjects difficult to consider in detail in the ESMO Clinical Practice Guidelines. The following areas were identified: (i) the elderly patient, (ii) prognostic factors suitable for clinical use and (iii) the ‘ultra-high-risk’ group. Before the conference, the expert panel was divided into three working groups; each group focused on one of these areas in order to address four clinically relevant questions relating to that topic. All relevant scientific literature, as identified by the experts, was reviewed in advance. During the consensus conference, each working group developed recommendations to address each of the four questions assigned to their group. These recommendations were then presented to the entire panel and a consensus was reached. This manuscript presents recommendations dedicated to the second area of interest, i.e. prognostic factors suitable for clinical use. The four topics [i.e. interim positron emission tomography (PET), TP53 mutations, cell of origin (COO) and minimal residual disease (MRD)] were primarily chosen because of the bulk of available data together with the lack of clear guidance regarding their use in clinical practice and within clinical trials. Results, including a summary of evidence supporting each recommendation, are detailed in this manuscript. The panel acknowledged that detection of TP53 inactivation by deletion or mutation in CLL should be implemented in clinical practice (level of evidence I, strength of recommendation A). Due to their potentially high prognostic value, at least in some lymphoma entities, implementation of interim PET, COO and MRD was highly recommended in the context of clinical trials. All expert panel members approved this final article.

Key words: lymphoma, consensus, positron emission tomography, TP53, cell of origin, minimal residual disease

introduction

Laboratory-based and imaging tools are increasingly used in patients with lymphoid malignancies to better understand their prognosis and even to guide therapeutic decisions. Despite their documented predictive value in several specific settings, their use is often extended to conditions where there is little evidence of substantial therapeutic benefit. This could result in an increase in costs and inappropriate therapeutic decisions. As such, clear recommendations regarding the use of these tools are required.

In 2015, the European Society for Medical Oncology (ESMO) held a consensus conference on mature B-cell neoplasms and chronic lymphocytic leukaemia (CLL) in order to develop recommendations on critical subjects that were difficult to consider in detail in the ESMO Clinical Practice Guidelines (CPG). In this consensus conference, one of the working groups (Working Group 2) focused on prognostic factors suitable for clinical use.
As such, the objectives of this working group were: (i) to identify a restricted number of prognostic tools whose clinical use is established or under rapid technological development; (ii) to discuss the technical and clinical reliability of these prognosticators; (iii) to consider the prognostic value of these tools; (iv) to provide recommendations on the use of these prognosticators in the context of clinical research and routine practice. Here, we describe the recommendations developed by Working Group 2 and approved by the whole panel, and provide a summary of evidence supporting each recommendation.

**Methods**

A consensus panel, comprising a multidisciplinary panel of 25 experts in the management of lymphoma, was convened by ESMO. Three consensus conference chairs (CB, ML, MH) were also appointed. The consensus panel was divided into three working groups, each of which was assigned a specific subject area and a working group chair as follows: Working Group 1: the elderly patient (Chair: CB); Working Group 2: prognostic factors suitable for clinical use (Chair: ML); Working Group 3: the ‘ultra-high-risk’ group (Chair: MH). The consensus conference was held on 20 June 2015 in Lugano, Switzerland. Before this consensus conference, four clinically relevant questions were identified for each subject area.

A literature review was conducted by each working group before the consensus conference, with each group responsible for compiling a summary of relevant information required to develop recommendations relating to each of their questions at the conference. No systematic literature search was undertaken. During the conference, in parallel sessions, the three working groups discussed and agreed on recommendations relating to each of their assigned questions. The level of evidence and strength of each recommendation were also noted, which were defined based on the 'Infectious Diseases Society of America-United States Public Health Service Grading System', as shown in Table 1 [1]. Recommendations from each group were then presented to the entire panel of experts, where they were discussed and modified, as required. Finally, a vote was conducted to determine the level of agreement among the expert panel for each of the recommendations. Discussion regarding each of the recommendations was completed after the consensus meeting, with additional supporting evidence published after the meeting also included in the final manuscript.

For Working Group 2, which is the focus of this report, four prognostic tools were identified for discussion in terms of their potential suitability as prognostic tools for clinical use. Discussions focused on B-cell lymphoma and CLL; plasma cell disorders and T-cell lymphoma were considered outside the scope of this consensus conference. In addition, working group members were asked to focus on disease entities in which the prognostic tools were most promising and where a greater need for clinical recommendations was required (front-runner entities, FRE). As such, the following prognostic tools and associated disease entities of specific interest were considered:

(i) interim positron emission tomography [PET; FRE: Hodgkin’s lymphoma (HL), diffuse large B-cell lymphoma (DLBCL)],
(ii) TP53 mutations and deletions (FRE: CLL),
(iii) cell of origin (COO) determination by gene expression profiling (GEP) or immunohistochemistry (IHC) (FRE: DLBCL),
(iv) molecular-based minimal residual disease (MRD) evaluation [FRE: mantle cell lymphoma (MCL); follicular lymphoma (FL), CLL]

Results from the section of the consensus conference dedicated to prognostic factors suitable for clinical use, together with a summary of evidence supporting each recommendation, are detailed in this article. A summary of these recommendations is shown in Table 2. Importantly, these additional recommendations should be read in conjunction with the already-published ESMO CPGs for the diagnosis, treatment and follow-up of malignant lymphomas and CLL [2–6].

**Results**

1. Interim PET as a prognostic tool

18-F-fluorodeoxyglucose (FDG)-PET has recently been recommended as the standard tool for the evaluation, staging and response assessment for patients with FDG-avid lymphomas, including HL, DLBCL and FL [7]. With the use of FDG-PET, metabolic response has increasingly been acknowledged as one of the strongest available prognostic tools and has been identified as a surrogate test for chemosensitivity. For the purposes of this consensus manuscript, the definition of interim PET applies to any FDG-PET carried out during a planned systemic treatment, usually after 2–4 cycles in the case of a conventional chemotherapy programme, or after 2–4 cycles of reinduction chemotherapy.
Table 2. Summary of recommendations

<table>
<thead>
<tr>
<th>Guidelines statement</th>
<th>LoE</th>
<th>GoR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. The potential role of interim PET as a prognostic tool</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recommendations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.1 The exploratory use of interim FDG-PET as a surrogate test of chemosensitivity and as a diagnostic tool to facilitate clinical decision-making is encouraged in clinical trials in HL, DLBCL and other aggressive FDG-avid lymphoma entities</td>
<td>III</td>
<td>B</td>
</tr>
<tr>
<td>1.2 There are little published data from randomised trials to support the use of an interim PET-driven therapeutic strategy in HL, DLBCL or other FDG-avid lymphomas. However, preliminary data strongly support the use of interim PET to tailor therapy in individual cases. On these grounds, results of interim PET may be applied in individual patients with early or advanced HL</td>
<td>II</td>
<td>C</td>
</tr>
<tr>
<td>1.3 Based on the lack of therapeutic consequences, the routine clinical use of interim PET is not recommended in patients with DLBCL</td>
<td>II</td>
<td>D</td>
</tr>
<tr>
<td>1.4 Based on the lack of data, the routine use of interim PET as a decision tool is discouraged in non-HL, non-DLBCL, FDG-avid lymphoma entities</td>
<td>V</td>
<td>E</td>
</tr>
<tr>
<td>2. The potential role of TP53 mutations and deletions as a prognostic tool</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recommendations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.1 Given the well-established, prognostic and predictive value of TP53 disruption in CLL, the panel strongly recommends the inclusion of TP53 analysis, both by FISH and DNA sequencing, in clinical trials of CLL for intervention and monitoring purposes. In particular, the availability of new drugs that overcome TP53-mediated chemoresistant disease mandates the acquisition of TP53 status for all patients with CLL at the time of screening procedures in trials in which one or more arms may be based on drugs that are known to be ineffective in TP53-disrupted CLL. Therefore, the use of TP53 screening for monitoring and intervention in clinical trials is encouraged in CLL</td>
<td>I</td>
<td>A</td>
</tr>
<tr>
<td>2.2 In other lymphoid neoplasms, TP53 screening for investigational purposes is neither recommended nor discouraged. At present, the panel discourages clinical trials aimed at specific interventions based on TP53 status unless prognostic markers are the major focus of the trial and the drug being evaluated has a strong biological rationale for overcoming TP53-mediated resistance</td>
<td>V</td>
<td>C (investigation); D (intervention)</td>
</tr>
<tr>
<td>2.3 In CLL, the panel supports analysis of TP53 disruption at the time of treatment requirement, both in first-line and subsequent lines of therapy. Reassessing TP53 status in previously TP53 wild-type CLL at relapse requiring treatment is relevant since TP53 disruption may develop, or become detectable only at relapse. In routine practice, characterising TP53 status in a given patient with CLL is clinically relevant as this may affect treatment decisions. The use of TP53 screening by FISH and mutational analysis for monitoring and intervention in clinical practice is therefore encouraged in CLL, provided there is availability of and access to therapies overcoming TP53-mediated resistance (e.g. inhibitors of the B-cell receptor and allo-SCT)</td>
<td>I</td>
<td>A</td>
</tr>
<tr>
<td>2.4 In other lymphoid neoplasms, the panel discourages the use of TP53 outside of clinical trials as there is no general recommendation for treatment modification currently published. The results of currently recruiting trials might modify this attitude in the coming years</td>
<td>V</td>
<td>E</td>
</tr>
<tr>
<td>3. The potential role of COO determination by IHC or GEP as a prognostic tool</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recommendations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.1 Given the limitations of IHC, the panel does not encourage its use in prospective clinical trials for prognostication</td>
<td>I</td>
<td>C</td>
</tr>
<tr>
<td>3.2 Given the limitations of IHC, the panel discourages its use in prospective clinical trials to guide intervention</td>
<td>I</td>
<td>D</td>
</tr>
<tr>
<td>3.3 The panel strongly encourages the use of GEP in prospective clinical trials for prognostication</td>
<td>I</td>
<td>A</td>
</tr>
<tr>
<td>3.4 Clinical trials of interventions based on GEP results are encouraged</td>
<td>I</td>
<td>B</td>
</tr>
<tr>
<td>3.5 Based on inadequate standardisation, and a lack of well-designed interventional studies, the use of COO determination by IHC or GEP in DLBCL is generally not recommended in routine clinical practice outside of clinical trials</td>
<td>V</td>
<td>D</td>
</tr>
<tr>
<td>4. The potential role of molecular-based MRD evaluation as a prognostic tool</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recommendations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.1 The use of MRD evaluation for monitoring and intervention in clinical trials is encouraged in MCL, FL and CLL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(i) MCL: for monitoring</td>
<td>I</td>
<td>B</td>
</tr>
<tr>
<td>(ii) MCL: for intervention</td>
<td>III</td>
<td>C</td>
</tr>
<tr>
<td>(iii) FL: for monitoring</td>
<td>I</td>
<td>B</td>
</tr>
<tr>
<td>(iv) FL: for intervention</td>
<td>IV</td>
<td>C</td>
</tr>
<tr>
<td>(v) CLL: for monitoring (depends on the drug used)</td>
<td>I</td>
<td>B</td>
</tr>
<tr>
<td>(vi) CLL: for intervention</td>
<td>IV</td>
<td>C</td>
</tr>
</tbody>
</table>
before the administration of a planned high-dose chemotherapy followed by stem cell support where intensified regimens are used.

In HL and DLBCL, the identification of metabolic response during treatment has been correlated with the individual risk of relapse, and of death in some cases, and has the potential to improve patient outcome through the early adaptation of treatment intensity [8–12]. There is general consensus that the achievement of an early metabolic response during treatment is predictive of favourable outcomes in terms of both progression and overall survival (OS). The high negative predictive value of interim PET, however, is counterbalanced by a variable rate of false-positive results that are usually more common in DLBCL than in HL [13].

methodological considerations

broad availability. Although FDG-PET is broadly available in high-income and in some middle-income countries, it remains inaccessible for many patients. As such, access to FDG-PET still needs to be improved worldwide.

reproducibility and standardisation. Reproducibility of FDG-PET has markedly improved with the application of standardised and recommended methods, particularly with the use of the Deauville 5-point scale (5PS) [14–17] [III, B]. However, quality assurance and training programmes are still needed. The application of semi-quantitative measurements of interim PET [i.e. delta standardised uptake value (SUV) max] is not recommended, although data suggest it may add prognostic detail in DLBCL [18].

clarity of reporting system. Currently, routine clinical reports are not well standardised. The panel recommends documenting the 5PS and SUV of the main lesions in the interim FDG-PET report of patients receiving front-line treatment [19].

prognostic value

The panel was confident of the high prognostic value of interim FDG-PET when used during induction therapy with anthracycline-containing regimens (after 2–4 cycles) in immunocompetent and HIV-negative patients with DLBCL [III, A] [11, 13, 16, 18, 26–29].

Finally, the panel recognised that interim FDG-PET is prognostic when used after reinduction chemotherapy and before a preplanned high-dose therapy programme in relapsed or refractory HL and DLBCL [III, A] [9, 10, 30–36].

panel recommendations for the use of interim PET for monitoring and intervention in clinical trials

recommendation 1.1. The exploratory use of interim FDG-PET as a surrogate test of chemosensitivity and as a diagnostic tool to facilitate clinical decision-making is encouraged in clinical trials in HL, DLBCL and other aggressive FDG-avid lymphoma entities.

Level of evidence: III
Strength of recommendation: B
Consensus: 100% yes (23 voters)

recommendation 1.2. There are little published data from randomised trials to support the use of an interim PET-driven therapeutic strategy in HL, DLBCL or other FDG-avid lymphomas. However, preliminary data strongly support the use of interim PET to tailor therapy in individual cases. On these grounds, results of interim PET may be applied in individual patients with early or advanced HL [20–22, 25, 37–40].

Level of evidence: II
Strength of recommendation: C
Consensus: 100% yes (23 voters)

recommendation 1.3. Based on the lack of therapeutic consequences, the routine clinical use of interim PET is not recommended in patients with DLBCL [13].

Level of evidence: II
Strength of recommendation: D
Consensus: 100% yes (23 voters)

recommendation 1.4. Based on the lack of data, the routine use of interim PET as a decision tool is discouraged in non-HL, non-DLBCL, FDG-avid lymphoma entities.

Level of evidence: V
Strength of recommendation: E
Consensus: 100% yes (23 voters)

Table 2. Continued

<table>
<thead>
<tr>
<th>Guidelines statement</th>
<th>LoE</th>
<th>GoR</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.2 The use of MRD evaluation for monitoring and intervention in clinical practice is not recommended in MCL, FL and CLL, with the exception of monitoring after allo-SCT</td>
<td>V</td>
<td>D</td>
</tr>
</tbody>
</table>

*There was 100% consensus from the panel of experts for all recommendations listed.

LOE, level of evidence; GOR, grade of recommendation; allo-SCT, allogeneic haematopoietic stem cell transplantation; FL, chronic lymphocytic leukaemia; COO, cell of origin; DLBCL, diffuse large B-cell lymphoma; DNA, deoxyribonucleic acid; FDG, fluorodeoxyglucose; FISH, fluorescence in situ hybridisation; FL, follicular lymphoma; GEP, gene expression profiling; HL, Hodgkin’s lymphoma; IHC, immunohistochemistry; MCL, mantle cell lymphoma; MRD, minimal residual disease; PET, positron emission tomography.
2. The potential role of TP53 mutations and deletions as a prognostic tool

The tumour suppressor gene TP53 maps at 17p13 and codes for a central regulator of the deoxyribonucleic acid (DNA) damage-response pathway; its activation leads to cell cycle arrest and DNA repair, apoptosis or senescence [41, 42]. In lymphoid response pathway; its activation leads to cell cycle arrest and a central regulator of the deoxyribonucleic acid (DNA) damage-

The tumour suppressor gene TP53 is the most frequently disrupted gene in lymphoid malignancies, and may progressively increase at the time of relapse or development of chemorefractory disease, as clearly documented in the case of CLL [44]. The fact that TP53 disruption may be acquired during the disease course is important from a diagnostic perspective, requiring, where clinically indicated, the sequential analysis of the locus at each time of treatment requirement [45–48]. The clinical importance of TP53 abnormalities in lymphoid malignancies is best demonstrated in the case of CLL, where TP53 disruption is tightly linked to the poor prognosis marked by this genetic lesion and its close association with chemorefractory disease, as documented by a number of observational studies and prospective trials conducted both in the chemotherapy and immuno-chemotherapy eras [49–55]. However, there is evidence that TP53 disruption predicts an adverse outcome also in other mature B-cell neoplasms [56, 57].

methodological considerations

broad availability. A complete analysis of TP53 disruption requires the availability of both FISH and DNA sequencing. Analysis of TP53 deletion by FISH is widely available in many haematological referral centres as well as in diagnostic laboratories dedicated to genetic disorders. Conversely, analysis of TP53 mutations by Sanger sequencing is currently restricted to highly specialised centres, and is not widely available. The panel agrees that, at least in the context of CLL, a complete analysis of TP53 disruption, including analysis of TP53 mutations, should be prioritised because TP53 disruption is the only well-established genetic marker which requires adaptation of treatment in CLL [45–48].

reproducibility and standardisation. FISH analysis for del17p13 is considered a well-standardised and reproducible technique. Sanger sequencing analysis for TP53 mutations is technically well standardised and adequately reproducible in experienced laboratories. Until recently, inter-laboratory reproducibility has not been systematically assessed. However, the European Research Initiative on CLL (ERIC) has now implemented a quality control initiative for TP53 mutations in many centres in Europe [58].

clarity of reporting system. Currently, there is no standardised reporting system for TP53 analysis across different centres. Data derived from randomised trials supporting these recommendations were obtained using a cut-off for FISH of 10%–20% of positive cells by Sanger sequencing. Regarding TP53 mutation analysis, the cut-off for mutation detection by Sanger sequencing can be generally estimated at 15%–20% of positive cells, although it may vary according to the precise nucleotide position and sequence. Inter-observer variability in the interpretation of electropherograms may also affect the detection threshold of Sanger sequencing; the use of dedicated software for mutation detection may reduce, at least in part, such variability. The precise description of TP53 mutations should be documented according to the well-codified Human Genome Variation Society (HGVS) nomenclature system (www.hgvs.org/mutnomen). Mutations also need to be validated through the International Agency for Research on Cancer (IARC) TP53 database (p53.iarc.fr). The GenBank reference sequence used for mutation detection should also be clearly stated in diagnostic reports.

prognostic value

The panel is confident with the general prognostic and predictive value of TP53 disruption in CLL [I, A]. The panel is also confident with the general prognostic value of TP53 disruption in other diseases, namely MCL, DLBCL and FL [II, B].

Many studies, both prospective and retrospective, have demonstrated that TP53 disruption is associated with a poor prognosis in CLL [48–55]. In particular, the CLL8 trial of the German CLL Study Group clearly documented that both del17p13 and TP53 mutation identify a very high-risk category of patients with CLL who were treated with fludarabine/cyclophosphamide/rituximab (FCR), an immuno-chemotherapy regimen that is the gold standard first-line treatment for fit patients with CLL [51, 55]. Notably, the poor prognosis associated with TP53 disruption in CLL appears to be independent of the chemotherapeutic agents utilised [48–55]. However, this might potentially change when non-genotoxic drugs, such as ibritinib, idelalisib and venetoclax, become part of routine practice.

panel recommendations for molecular and cytogenetic analysis of TP53 disruption in CLL and other lymphoid neoplasms for monitoring and intervention in clinical trials

recommendation 2.1. Given the well-established, prognostic and predictive value of TP53 disruption in CLL, the panel strongly recommends the inclusion of TP53 analysis, both by FISH and DNA sequencing, in clinical trials of CLL for intervention and monitoring purposes. In particular, the availability of new drugs that overcome TP53-mediated chemorefractory disease mandates the acquisition of TP53 status for all patients with CLL at the time of screening procedures in trials in which one or more arms may be based on drugs that are known to be ineffective in TP53-disrupted CLL [58–62]. Therefore, the use of TP53 screening before the start of treatment is highly encouraged in CLL.

Level of evidence: I
Strength of recommendation: A
Consensus: 100% yes (23 voters)

recommendation 2.2. In other lymphoid neoplasms, TP53 screening for investigational purposes is neither recommended nor discouraged. At present, the panel discourages clinical trials aimed at specific interventions based on TP53 status unless prognostic markers are the major focus of the trial and the drug being
evaluated has a strong biological rationale for overcoming TP53-mediated resistance.

Level of evidence: V
Strength of recommendation for investigation: C
Strength of recommendation for intervention: D
Consensus: 100% yes (23 voters)

panel recommendations for molecular and cytogenetic analysis of TP53 disruption in CLL and other lymphoid neoplasms for monitoring and intervention in clinical practice outside of clinical trials

**recommendation 2.3.** In CLL, the panel supports analysis of TP53 disruption at the time of treatment requirement, both in first-line and subsequent lines of therapy. Reassessing TP53 status in previously TP53 wild-type CLL at relapse requiring treatment is relevant since TP53 disruption may develop, or become detectable only at relapse. In routine practice, characterising TP53 status in a given patient with CLL is clinically relevant as this may affect treatment decisions. The use of TP53 screening by FISH and mutational analysis for monitoring and intervention in clinical practice is therefore encouraged in CLL, provided there is availability of and access to therapies overcoming TP53-mediated resistance [e.g. inhibitors of the B-cell receptor and allogeneic haematopoietic stem cell transplantation (allo-SCT)].

Level of evidence: I
Strength of recommendation: A
Consensus: 100% yes (23 voters)

**recommendation 2.4.** In other lymphoid neoplasms, the panel discourages the use of TP53 outside of clinical trials as there is no general recommendation for treatment modification currently published. The results of currently recruiting trials might modify this attitude in the coming years.

Level of evidence: V
Strength of recommendation: E
Consensus: 100% yes (23 voters)

### 3. The potential role of COO determination by IHC or GEP as a prognostic tool

DLBCL is the most common form of lymphoma in the Western world [63]. It shows a wide spectrum of morphology and is biologically heterogeneous [63]. To identify biological entities within DLBCL, GEP has been applied to tumour samples of DLBCL [64–66]. The seminal study by Alizadeh et al. [64] was the first to recognise that DLBCL contains at least two biological entities, one with a GEP similar to the normal purified germinal centre B-cell (the germinal centre B-cell profile, or GCB) and the other similar to the profile produced by a purified, in vitro immunoglobulin M (IgM)-stimulated B-cell (the activated B-cell profile, or ABC). Consequently, DLBCL was commonly divided into these two subtypes, which show different clinical and molecular features. The robustness of this profile based on GEP has been confirmed in other studies [67–69].

As the use of high-throughput GEP was considered unfeasible in routine laboratory practice, there have been several attempts to simplify the procedures for COO determination. These attempts have gone in two directions, namely the identification of IHC surrogates and the application of GEP (either high-throughput or low-throughput) to formalin-fixed and paraffin-embedded (FFPE) samples.

Several IHC surrogate protocols use an algorithm to identify the COO in FFPE samples of DLBCL. Several algorithms have been published, including those by Colomo et al. [70], Hans et al. [71], Muris et al. [72], Choi et al. [73], Nyman et al. [74], Natkunam et al. [75], Meyer (better known as ‘Tally’) et al. [76] and Visco et al. [77]. Although these seem to work well as survival predictors when samples are stained and analysed in a single centre, the results are not easily transferable to other laboratories [78–80]. Indeed, data from a large randomised clinical trial [81] and a meta-analysis have shown a limited role for IHC algorithms [82].

Given the limitations of IHC in terms of COO signature reproducibility, several groups have attempted to use FFPE as a source of RNA to identify the COO signature by high-throughput [83, 84] or low- to medium-throughput GEP techniques [85–91]. The results have been much more robust than those obtained by IHC, and findings from a meta-analysis have confirmed the usefulness of GEP approaches [82]. Most studies were retrospective, but two phase III clinical trials incorporating COO GEP on FFPE samples, namely the REMoDL-B study (ClinicalTrials.gov identifier NCT01324596 [92]), which uses the Illumina DASL platform (Illumina Inc., San Diego, CA), and the ROBUST study (ClinicalTrials.gov identifier NCT02285062 [93]), which uses the Nanostring nCounter-based Lymph2Cx platform (NanoString Technologies, Seattle, WA), are ongoing [94].

Against this background, the panel members discussed the adequacy for clinical use of COO-determining methods in DLBCL by both IHC and GEP.

**methodological considerations**

**broad availability.** IHC is widely available. Conversely, GEP technologies are currently limited to very specialised laboratories. The introduction of more user-friendly technologies (such as Nanostring nCounter) might render GEP more widely available and applicable in routine clinical practice in the near future.

**reproducibility.** IHC suffers from major reproducibility issues, which include inter-laboratory and inter-observer concordance, varying degrees of overlap with the gold standard GEP techniques and often poor correlation between the various algorithms available [78–82]. GEP using well-established high-throughput commercial chips is robust; however, inter-laboratory variability needs to be explored. So far, only one study using the Nanostring-based Lymph2Cx assay has assessed inter-laboratory agreement, with excellent results; the same test also showed excellent concordance for resampled biopsies and between different reagent lots [88, 91]. However, processing of samples would critically influence the outcome of GEP results and so particular care should be devoted to pre-analytical variables.

**clarity of reporting systems.** The reporting system for IHC has been standardised, with algorithms to clearly specify thresholds and procedures (e.g. the Hans classifier uses a 30% positive cell cut-off and a step-by-step algorithm), although few pathology reports specify the exact percentage of positive cells or even the algorithm used. For GEP, no standardised system for the interpretation of data or reporting of results is available.
prognostic value

The limitations of IHC algorithms have been described earlier. The panel also raised substantial concerns regarding the prognostic value of IHC. Conversely, several published studies support the general prognostic value of COO assessment by GEP in DLBCL, and so the panel was more confident in supporting this technical approach for prognostication [I, A] [82]. These considerations are particularly relevant with regard to drugs, which promise differential activity in germinal centre B-cell-like versus activated B-cell-like DLBCLs.

panel recommendations for the use of COO identification by IHC and GEP in clinical trials for monitoring and intervention

**recommendation 3.1.** Given the limitations of IHC, the panel does not encourage its use in prospective clinical trials for prognostication.

- Level of evidence: I
- Strength of recommendation: C
- Consensus: 100% yes (23 voters)

**recommendation 3.2.** Given the limitations of IHC, the panel discourages its use in prospective clinical trials to guide intervention.

- Level of evidence: I
- Strength of recommendation: D
- Consensus: 100% yes (23 voters)

**recommendation 3.3.** The panel strongly encourages the use of GEP in prospective clinical trials for prognostication.

- Level of evidence: I
- Strength of recommendation: A
- Consensus: 100% yes (23 voters)

**recommendation 3.4.** Clinical trials of interventions based on GEP results are encouraged.

- Level of evidence: I
- Strength of recommendation: B
- Consensus: 100% yes (23 voters)

panel recommendations for the use of COO identification by IHC and GEP for monitoring and intervention in routine clinical practice

**recommendation 3.5.** Based on inadequate standardisation, and a lack of well-designed interventional studies, the use of COO determination by IHC or GEP in DLBCL is generally not recommended in routine clinical practice outside of clinical trials.

- Level of evidence: V
- Strength of recommendation: D
- Consensus: 100% yes (23 voters)

4. MRD evaluation by polymerase chain reaction-based methods and flow cytometry

MRD assessment can be used for the identification of different prognostic subgroups in patients with B-cell lymphomas and CLL, and is an excellent surrogate for treatment outcome [95–99]. Published evidence for the prognostic impact of MRD exists for MCL [96, 100, 101], FL [95, 96, 98, 102, 103] and CLL [97, 104–107]. In these entities, achievement of MRD response by conventional or intensified treatment is associated with prolonged progression-free survival (PFS) and OS independent of categorical response assessment and a favourable prognosis. Several prospective phase III trials using standardised approaches for MRD assessment have been published and demonstrate the prognostic relevance of MRD response in FL, MCL and CLL independently of treatment regimen or strategy and clinical risk parameters [95–99]. Indeed, the prognostic impact of MRD status has led to MRD being proposed as a secondary end point in ongoing clinical trials. In CLL, recent evidence suggests that MRD might also be used to identify candidates for dose de-escalations. Therefore, polymerase chain reaction (PCR)-based MRD evaluation is considered a promising prognosticator in MCL and FL, whereas MRD evaluation by flow cytometry is preferred in CLL.

methodological considerations

**broad availability.** Flow cytometry is generally available in Europe for CLL, but standardised four-colour flow to detect MRD at a level of 10–4 is only available in specialised institutions; real-time quantitative (RQ)-PCR is only available in specialised centres (EURO MRD network; www.euromrd.org).

**reproducibility and standardisation.** For RQ-PCR, reproducibility is excellent and methods are standardised and subjected to periodic quality controls at specialised institutions involved in the EURO MRD network. Flow-based MRD methods are currently harmonised, but not standardised, and inter-laboratory reproducibility has not been systematically assessed.

**clarity of reporting systems.** Reporting of molecular MRD results is standardised within established networks. Flow cytometry standardisation is currently ongoing within the EuroFlow network (http://www.euroflow.org).

prognostic value

The panel is confident of the general prognostic value of MRD evaluation in MCL [I, A], FL [I, A] and CLL [I, A].

Several phase III clinical trials have been carried out in FL [95, 102, 108, 109], CLL [97, 110, 111] and MCL [96] that clearly demonstrate the usefulness of MRD as a surrogate end point for monitoring treatment efficiency and for its prognostic value. Remarkably, in all three entities, the prognostic impact of MRD response on PFS and OS has been documented independent of treatment regimen, mostly in both peripheral blood and bone marrow.

panel recommendations for the use of MRD evaluation in clinical trials for monitoring and intervention

The panel felt confident that MRD monitoring of treatment response as an end point in clinical trials might facilitate the interpretation of results. Whether trials investigating MRD-based treatment tailoring might lead to substantial therapeutic improvement and treatment optimisation is an attractive but as yet unproven possibility. MRD assessment post-induction therapy is the most frequently assessed time point for MRD response as it is associated with a high prognostic impact and is therefore suitable to guide treatment intervention. Later time points during treatment are also of prognostic value and are suitable to
guide treatment intervention. However, so far, only one clinical trial in CLL has been published, which showed that MRD-based intervention (in terms of discontinuation of treatment once MRD negativity was seen) was associated with comparable PFS and OS independent of the number of courses of treatment received [110]. The panel therefore decided that more data are required to support the clinical benefit of treatment modification based on efficacy, as determined by MRD negativity.

**recommendation 4.1.** The use of MRD evaluation for monitoring and intervention in clinical trials is encouraged in MCL, FL and CLL:

(i) MCL: for monitoring [112, 113];
   - Level of evidence: I
   - Strength of recommendation: B

(ii) MCL: for intervention [112, 113];
   - Level of evidence: III
   - Strength of recommendation: C

(iii) FL: for monitoring [113];
   - Level of evidence: I
   - Strength of recommendation: B

(iv) FL: for intervention [113];
   - Level of evidence: IV
   - Strength of recommendation: C

(v) CLL: for monitoring (depends on the drug used):
   - Level of evidence: I
   - Strength of recommendation: B

(vi) CLL: for intervention:
   - Level of evidence: IV
   - Strength of recommendation: C

Consensus: 100% yes (23 voters)

**Panel recommendations for the use of MRD for monitoring and intervention in routine clinical practice**

The panel does not support MRD evaluation for monitoring or intervention in routine practice outside of clinical trials as there is no general recommendation for treatment modification currently published. However, results of ongoing trials might modify this attitude in the coming years. The only exception is MRD assessment after allo-SCT, where it is a useful tool to monitor lymphoma regrowth and is more sensitive than currently used short tandem repeat analysis. In this setting, MRD can be used for discontinuation or intensification of immunosuppression [114].

**recommendation 4.2.** The use of MRD evaluation for monitoring and intervention in clinical practice is not recommended in MCL, FL and CLL, with the exception of monitoring after allo-SCT.

- Level of evidence: V
- Strength of recommendation: D
- Consensus: 100% yes (23 voters)

**Acknowledgements**

The authors thank Jennifer Lamarre, Claire Bramley, Matthew Wallace, Aude Galli and all ESMO staff for their support throughout the whole consensus process. Angela Corstorphine of Kstorfin Medical Communications Ltd provided medical writing support with the preparation of this manuscript. This support was funded by ESMO.

**Funding**

All costs relating to the consensus conference were covered from the European Society for Medical Oncology central funds. There was no external funding of the event or manuscript production.

**Disclosure**

CB (honoraria from Roche, Pfizer, Celgene, Pharmacyclics and Janssen; research grants from Roche and Janssen), MD (advisory boards for Bayer, Celgene, Gilead, Janssen and Pfizer; support for academic trials from Celgene, Janssen, Mundipharma, Pfizer and Roche; speaker’s honoraria from Celgene, Janssen, Mundipharma, Pfizer and Roche), AJMF (advisory boards for Celgene, Gilead and Mundipharma; speaker honoraria from Italfarmaco, Mundipharma and Gilead; research support from Celgene, Roche, Mundipharma and Rhizen), GG (advisory boards for Roche, Janssen, Amgen, Novartis, GlaxoSmithKline, Karyopharm and Morphosys; speaker honoraria from Roche, Janssen, Gilead and GlaxoSmithKline), VG (speaker honoraria, advisory boards, consultancy and travel grants from Roche; speaker honoraria from Mundipharma, GlaxoSmithKline and Bristol-Myers Squibb), MH (speaker honoraria, advisory boards and consultancy from Takeda; advisory boards and consultancy from Janssen, Celgene and Roche/Genentech), ML (honoraria from Celgene, Janssen-Cilag, Roche, Amgen, Mundipharma and Teva; research contracts with Celgene, Pfizer, Mundipharma and Roche; research grants from Amgen, Roche and Takeda), SLG (speaker honoraria, advisory boards, consultancy and travel grants from Roche, Janssen-Cilag and Celgene; research programme funding from Roche, Janssen and Servier), UM (honoraria for advisory board activities from Roche, Celgene, Amgen and Janssen-Cilag; research grants from Mundipharma and Celgene), PdNB (speaker’s bureau/lectures for Roche and Bayer), MP (advisory boards for Celgene, Roche and Spectrum; research support from Amgen, Roche and Spectrum), NS (travel grants, advisory boards and research support from Roche; travel grants and advisory boards from Celgene; travel grants and honoraria for company symposia from Riemsner Pharma and CTI Life Sciences), PS (advisory boards for Celgene and Teva; honoraria from Spectrum Pharmaceuticals and Pierre Fabre Oncology; travel grants from Hospira, Celgene and Teva), MS (speaker’s bureau and advisory boards for Teva, Mundipharma, Janssen-Cilag, Gilead, CTI BioPharma and Servier; research support, consultant, speaker’s bureau and advisory boards for Menarini; speaker’s bureau for Roche and Takeda), RS (research support from Celgene, Novartis and Teva; honoraria from Celgene, Novartis and Teva; scientific advisory board for Celgene), ASB (consultancy fees from Takeda), MT (consultancy for Roche, Gilead, Janssen, Takeda and Celgene), JW (advisory boards for Roche, Celgene, Janssen-Cilag and Takeda; research grants from Roche, Celgene, Janssen-Cilag, Takeda, GlaxoSmithKline and Seattle Genetics), UW (speaker honoraria from Amgen, Novartis, Roche and Chugai), EZ (advisory honoraria and/or support for investigator-initiated studies for the
institution from Celgene, Johnson and Johnson/Janssen, Gilead, Mundipharma, Roche and Bayer). All remaining authors have declared no conflicts of interest.

references


appendix

ESMO Lymphoma Consensus Conference Panel Members

C Buske, Comprehensive Cancer Center Ulm and Department of Internal Medicine III, Institute of Experimental Cancer Research, University Hospital, Ulm, Germany; M Dreyling, Medizinische Klinik III, Klinikum der Universität München/ LMU, Munich, Germany; AJM Ferreri, Unit of Lymphoid Malignancies, Department of Onco-Hematology, IRCCS San Raffaele Scientific Institute, Milan, Italy; P Fields, Department of Haematology, Guys and St Thomas’ and King’s College Hospitals, London, UK; G Gaidano, Division of Hematology, Department of Translational Medicine, Amedeo Avogadro University of Eastern Piedmont, Novara, Italy; V Goede, Department of Internal Medicine, University Hospital Cologne, Germany; M Hutchings, Department of Hematology, Rigshospitalet, Copenhagen, Denmark; M Ladetto, Hematology Division, Azienda Ospedaliera Santi Antonio e Biagio e Cesare Arigo, Alessandria, Italy; SL Gouill, Centre Hospitalo-Universitaire de Nantes, UMR892 team 10, CIC Nantes, France; S Luminari, Hematology, Arcispedale S. Maria Nuova, IRCCS Reggio Emilia, and Department of Diagnostic, Clinical and Public Health Medicine, University of Modena and Reggio Emilia, Modena, Italy; U Mey, Department of Oncology and Haematology, Kantonsspital Graubünden, Chur, Switzerland; PDN Brown, Department of Hematology, Rigshospitalet, Copenhagen, Denmark; M Pfundschuh, Innere Medizin I, University Klinik des Saarlandes, Hamburg, Germany; C Pott, Second Medical Department, University Hospital Schleswig-Holstein, Kiel, Germany; N Schmitz, Department of Hematology, Oncology and Stem Cell Transplantation, Asklepios Klinik St Georg, Hamburg, Germany; P Soubyryan,
With the genomic revolution and the era of targeted therapy, prognostic and predictive gene signatures are becoming increasingly important in clinical research. They are expected to assist prognosis assessment and therapeutic decision making. Notwithstanding, an evidence-based approach is needed to bring gene signatures from the laboratory to clinical practice. In early breast cancer, multiple prognostic gene signatures are commercially available without having formally reached the highest levels of evidence-based criteria. We discuss specific concepts for developing and validating a prognostic signature and illustrate them with contemporary examples in breast cancer. When a prognostic signature has not been developed for predicting the magnitude of relative treatment benefit through an interaction effect, it may be wishful thinking to test its predictive value. We propose that new gene signatures be built specifically for predicting treatment effects for future patients and outline an approach for this using a cross-validation scheme in a standard phase III trial. Replication in an independent trial remains essential.

**Key words:** gene signature, prognostic, predictive, evidence based, clinical utility

### introduction

Molecular signatures are becoming increasingly important for anticipating the prognosis of individual patients (‘prognostic’ biomarkers) or for predicting how individual patients will respond to specific treatments (‘predictive’ biomarkers, more generally called ‘treatment-effect modifiers’). A voluminous literature of >150 000 papers documenting thousands of claimed biomarkers has been produced in medicine of which fewer than 100 have been validated for routine clinical practice [1]. Indeed, <20 prognostic or predictive biomarkers are recognized with variable levels of evidence in the 2014 European Society of Medical Oncology (ESMO) clinical practice guidelines for lung, breast, colon and prostate cancer [2].

In early breast cancer, while several clinical prediction models exist based on clinical and pathological (CP) characteristics, such as age, tumor size, nodal status, tumor grade, estrogen receptor, at least six different gene signatures are commercially available (OncoType DX, MammaPrint, Genomic Grade Index, PAM50, Breast Cancer Index and EndoPredict). The concordance of predicted risk categories of the different gene signatures for individual patients is moderate [3, 4], as illustrated by recent OPTIMA study which evaluated—among others—the two well-known tests Mammaprint (low/high) and OncoType DX (≤25 versus >25) on 302 patients in a head-to-head comparison and found a low level of agreement, i.e. a kappa value of 0.40 (95% CI 0.30–0.49) [5]. Of course, even when repeating the same assay twice on a single tumor sample, some inherent degree of inaccuracy would be expected but unlikely to this extent. This has led to a pretty awkward situation where the treatment decision for adjuvant chemotherapy does not depend anymore on the clinician but on the