

SUMMARY OF DISCUSSION



Cancer Drug Development Forum (CDDF) Multi-Stakeholder Workshop

Minimal Residual Disease

18–19 October 2017

London, UK

Prepared by Excerpta Medica

TABLE OF CONTENTS

PROGRAMME.....	3
BACKGROUND AND OBJECTIVES.....	5
THE REGULATOR'S VIEW: SURROGATE ENDPOINTS AND MRD	5
Definition of endpoints.....	5
The CHMP/EMA view.....	5
How to apply MRD as an intermediate endpoint in RCT?	6
Measuring MRD.....	6
Areas of uncertainty	7
The FDA perspective	7
US regulatory considerations	7
Statistical evaluation of a surrogate.....	8
MRD data in FDA applications	9
Areas of uncertainty	9
Conclusions.....	9
OVERVIEW OF VALIDATED METHODS AND THRESHOLDS USED TO ASSESS MRD IN MM	10
MRD diagnostics and CTCs: ELDA PCR assay in GMMG MM5 study.....	10
What is the optimal threshold?	10
Flow cytometry.....	12
Next-generation sequencing	12
Discussion	13
CURRENTLY AVAILABLE DATA: HOW TO DEFINE THE OPTIMAL DATASET FOR VALIDATION?	14
Overview of 2 published meta-analyses	14
PETHEMA/GEM Flow-MRD datasets	14
Myeloma MRD: UK data	15
In-depth look at specific datasets; which ones meet requirements? Individual data "owners"/cooperative groups.....	16
Discussion	16
CLINICAL DEVELOPMENT STRATEGY FOR VALIDATION OF MRD AS A SURROGATE ENDPOINT.....	18
Quantifying surrogacy.....	18
Discussion	19
MRD AS A SURROGATE ENDPOINT: INSIGHTS FROM CLL	20
Ensure broad access to techniques for measuring MRD and identifying suitable approaches to reporting response	20
Understand compartment effects	21
Improve knowledge of MRD kinetics and biological responses to optimize treatment	21
SUMMING UP AND KEY CONCLUSIONS	21
ABBREVIATIONS	23

PROGRAMME

Day 1

SESSION 1: SURROGATE ENDPOINTS AND MRD

Chairs: *Beatriz Flores and Mohamed Zaki*

13:40 **Overview of requirements for surrogate endpoint adoption by CHMP/EMA**

- Regulatory perspective – *Beatriz Flores*
- ASO-PCR based MRD quantification in NDMM – *Stefanie Huhn*

14:10 **Overview of validated methods and thresholds used to assess MRD in MM**

Andy Rawstron, Bruno Paiva, Hervé Avet-Loiseau

14:55 **Roundtable discussion**

15:25 **Coffee break**

SESSION 2: CURRENTLY AVAILABLE DATA: HOW TO DEFINE THE OPTIMAL DATASET FOR VALIDATION?

Chairs: *Mélanie Frigault and John Smyth*

15:55 **Overview of 2 published meta-analyses**

Walter Gregory

16:15 **In-depth look at specific datasets; which ones meet requirements? Individual data “owners”/cooperative groups**

Roger Owen, Bruno Paiva, Stefania Oliva

17:05 **Roundtable discussion**

Day 2

SESSION 3: THE REGULATOR'S VIEW

Chairs: *Davy Chiodin and Kyriaki Tzogani*

9:00 **The FDA view**

Nicole Gormley

9:30 **Clinical development strategy for validation of MRD as a surrogate endpoint**

Walter Gregory

10:00 **Roundtable discussion**

10:30 **Coffee break**

SESSION 4: MRD AS A SURROGATE ENDPOINT: WHAT IS BEYOND MM?

Chairs: *Davy Chiodin and Kyriaki Tzogani*

10:50 **MRD as an endpoint: insights from CLL**

Andy Rawstron

SESSION 5: WRAP UP SESSION

11:20 **Summary and discussions on next steps**

Facilitator: *John Smyth*

12:20 **End of workshop**

BACKGROUND AND OBJECTIVES

The multiple myeloma (MM) treatment landscape has changed drastically over recent years with the introduction of new classes of agents associated with significant improvements in response rates and depth of responses. This has created a need for more advanced, or additional, response evaluation methods that allow regulatory assessment of investigational compounds in an earlier timeframe than the traditional endpoints of progression-free survival (PFS) and overall survival (OS).

The objective of the workshop was to establish how minimal residual disease (MRD) negativity might be adopted as a surrogate or intermediate endpoint for PFS and/or OS in clinical trials.

THE REGULATOR'S VIEW: SURROGATE ENDPOINTS AND MRD

Definition of endpoints

Clinical benefit, which is used to support regular approval, could be summarized as the way a patient feels, functions, or survives (e.g. PFS, OS). A surrogate endpoint predicts clinical benefit, but is not itself a measure of clinical benefit. For example, PFS and OS are measures of clinical benefit (true/final endpoint). In the ideal situation, the surrogate is within the direct causal pathway of the disease and its impact on the ultimate clinical endpoint. Any treatment effect on the surrogate/biomarker should ideally fully capture the full treatment effect on the true clinical endpoint.

The CHMP/EMA view

Beatriz Flores, Medicines and Healthcare Products Regulatory Agency (MHRA), UK

Dr Flores provided an overview of the requirements for surrogate endpoint adoption by the Committee for Medicinal Products for Human Use (CHMP) in expectation of published guidance. These include: 1) prospective validation in randomized controlled trial(s) (RCT[s]) specific to the treatment and indication; 2) consistent association; 3) biological plausibility. She argued that the available data has shown prognostic impact, but not correlation between MRD and outcome. An additional limitation is that the available data are based on heterogeneous datasets in terms of study population, different treatments, transplant

eligibility, different assays, different MRD cut-off, etc. Moreover, it is unclear what differences in MRD levels between treatment arms in an RCT can *accurately* predict differences in PFS/OS. She mentioned that this would likely differ per setting, possibly driven by different residual disease biology, i.e. newly diagnosed MM (NDMM) versus relapsed/refractory MM (RRMM). Because of these limitations, MRD cannot be accepted as a surrogate endpoint; however, the European Medicines Agency (EMA) will consider it as an intermediate endpoint at the moment.

An **intermediate endpoint** is a measure of a treatment effect that is reasonably likely to predict clinical benefit. MRD as intermediate endpoint will be defined as an undetectable MRD Response Rate (RR) endpoint, *e.g. the proportion of ITT patients who achieve complete remission (CR) and undetectable MRD in bone marrow (BM) at a pre-specified time point after treatment*. The EMA will require confirmatory PFS/OS analysis within an agreed timeframe.

How to apply MRD as an intermediate endpoint in RCT?

The RCT should be powered for PFS as the primary endpoint and in certain situations may also need to be powered for OS. The statistical plan should pre-specify a difference in MRD between the arms post-treatment sufficient enough to predict a meaningful benefit in PFS/OS. All patients should be followed up for OS.

Recommended secondary endpoints include: sustained undetectable MRD (e.g. for ≥ 1 year in CR patients with normal imaging); disease-free survival (DFS) (from start of undetectable MRD until reappearance of detectable MRD). Recommended exploratory endpoints are: different MRD cut-offs, MRD in partial response (PR)/very good partial response (VGPR)/near complete remission (nCR), assessment of MRD in peripheral blood (PB), and MRD at different time points.

Regulatory approvals are based on risk/benefit assessments; therefore, a safety endpoint should always be considered if the goal is early approval based on MRD.

The EMA/283093/2016 Guideline on the use of MRD as a clinical endpoint in MM studies is in draft stage.

Measuring MRD

- Cut-off: < 1 in 10^5 residual tumour cells as recommended by the International Working Group (IWG)

- Sample: BM
- Any standardized methods such as next-generation sequencing (NGS)/flow cytometry (FC) will be accepted
- It is required that MRD is assessed after each treatment stage (induction, consolidation, ASCT, maintenance)

Areas of uncertainty

- What is the significance of achieving undetectable MRD earlier versus later?
- What is the best duration of treatments?
- Is there a need to use 2 different validated assays? Currently, this is a recommendation, not a requirement.
- Should we limit the endpoint to patients in CR?
- How should extramedullary disease be handled?
- Should the MRD endpoint include normal free light chain ratio (as a composite endpoint)? Currently, the regulators are inclined not to.
- What is the impact of immune-based therapies on MRD?

The FDA perspective

Nicole Gormley, Clinical Team Leader, Division of Hematology Products, Food and Drug Administration (FDA), USA

US regulatory considerations

Dr Gormley started off her talk by highlighting the multiple potential uses for MRD assessment. It is generally recognized as a prognostic biomarker. Clinical uses can include monitoring of relapse, guiding therapeutic decisions, and potentially discontinuation at some point. In terms of regulatory uses, at this stage, MRD could be used as a patient stratification factor; however, there is some uncertainty about using it as a patient selection factor, or for risk-based treatment assignment. MRD is not at the point yet of being accepted as a surrogate endpoint in MM.

In the USA there are 2 regulatory pathways for approval of a new therapeutic: regular approval, and accelerated approval for therapies of serious or life-threatening illness. Both of these approvals can be based on a surrogate endpoint, and either on an established surrogate with regular approval or one that is reasonably likely to predict clinical benefit in the case of accelerated approval. Accelerated approval may require post-approval trials to verify the anticipated clinical benefit. There are several ways by which surrogate endpoints can be qualified or accepted by the FDA:

- FDA's Biomarker Qualification Program (biomarkers could be used in multiple drug development programmes)
- Within a specific drug development programme (limitation: it is specific to the treatment, and not necessarily as easily extrapolated to other drug programmes or disease settings)
- Collaborative group interactions with the Agency

Regardless of the approach, a key consideration is the risk introduced by the biomarker, i.e. selection of inappropriate treatment for the particular patient. Similar to the EMA, the FDA also requires a biological rationale and understanding of the disease pathway. Additional considerations are the availability of validated assays and reproducibility of the data.

Statistical evaluation of a surrogate

The surrogacy can be evaluated at both the individual and trial level using meta-analytic approaches. Dr Gormley reviewed the [FLASH group meta-analysis](#), which established a CR rate at 30 months as a surrogate endpoint in follicular lymphoma (FL).

This endpoint was chosen as this was the time at which most patients have completed any additional induction treatment and maintenance for 2 years. The investigators conducted a MEDLINE database search for RCTs in previously untreated FL patients, and the primary surrogacy candidate, CR at 30 months, was evaluated on both the trial level and individual-patient-level surrogacy. There were 13 clinical trials and individual data of 3,837 patients. Individual-patient-level surrogacy was established in a Cox model, using a landmark approach. CR at 30 months was associated with improved PFS (HR 0.70; $p < 0.0001$). Trial-level surrogacy was demonstrated with 2 R^2 methodologies, meeting the pre-specified statistical criteria (R^2 WLS 0.88, 95% CI 0.77–0.96; R^2 Copula 0.86, 95% CI 0.72–1.00). Lastly, the surrogate threshold effect was also evaluated. This is the minimum treatment effect on a

surrogate that is necessary to predict a non-zero difference in PFS and it was determined as an odds ratio of at least 1.56.

The FLASH group effort is an example of collaboration between academia and industry; the group sought early guidance from the FDA.

MRD data in FDA applications

A review of internal databases between 2014 and 2016 disclosed that 13 of 34 applications featured MRD data across haematological–oncological indications (chronic lymphocytic leukaemia [CLL], MM, acute lymphoblastic leukaemia [ALL], acute myeloid leukaemia [AML]). Data were considered inadequate in 34% of these applications. The reasons for exclusions included: missing data, inconsistent testing across sample sources, high amounts of test failure rate, incomplete test characteristics data (i.e. limit of detection [LOD]), and incomplete planned statistical analysis.

Areas of uncertainty

- What rate of MRD improvement is clinically meaningful?
- Which level of MRD best correlates with PFS or OS?
- Is a method-agnostic approach acceptable (provided that the method used is analytically sound and the results are reproducible)? Is the assay applicable to all disease settings, i.e. RRMM?
- What is the surrogate threshold effect? What would be the clinical benefit difference that would be appropriate to statistically power the trial?
- What is the appropriate timing of the assessment per disease setting?
- How should extramedullary disease be handled?
- How should missing data be handled?
- Can it be applied to all disease settings? Should cytogenetics be included as risk stratification factors in trials?

Conclusions

Dr Gormley concluded that MRD has the potential to be a useful clinical and regulatory tool in MM. However, data collection and assay performance characteristics should be of significant rigour and completeness to allow for comprehensive assessments.

The FDA encourages pharmaceutical companies to seek scientific advice and guidance from regulators early on if they are considering the use of a surrogate endpoint in clinical trials. The agency is committed to working with the community on the development of MRD in haematological malignancies.

OVERVIEW OF VALIDATED METHODS AND THRESHOLDS USED TO ASSESS MRD IN MM

MRD diagnostics and CTCs: ELDA PCR assay in GMMG MM5 study

Stefanie Huhn, Molekularbiologisches Labor, Sektion Multiples Myelom, Medizinische Klinik V, Universitätsklinikum Heidelberg, Germany

Dr Huhn talked about PCR-based quantification of MRD in NDMM patients based on extreme limiting dilution analysis (ELDA), reaching a median sensitivity of 10^{-6} . The assay was used to assess MRD in BM and PB samples of patients with NDMM enrolled in the GMMG MM5 trial. The presented analysis evaluated whether the presence of circulating tumour cells (CTCs) in PB could serve as a surrogate of MRD in BM. Samples were collected longitudinally until the end of the study. Correlation between CTCs and tumour load in BM was assessed in patients with CR or suspected CR (n = 107).

A correlation was observed for patients with CTCs in PB and tumour load in BM, meaning that if there were tumour cells in PB, there were also tumour cells in BM. Measuring CTCs at diagnosis, PR, VGPR, and nCR revealed an increasing proportion of patients with no tumour cells in PB. This is prognostically relevant, as patients who are PB CTC-negative post-transplant have improved OS (HR 5.5; p = 0.023). Dr Huhn concluded that the data seem to suggest that BM biopsies could be avoided as long as there are no detectable CTCs in PB, and this could be applied to a “minimally invasive” diagnostic workup. However, the impact of International Staging System (ISS) stage, cytogenetic risk, and immunotherapy on CTCs as well as the biology of the residual disease would require further elucidation.

What is the optimal threshold?

Andy Rawstron, HMDS, St James's Institute of Oncology, Leeds Teaching Hospitals NHS Trust, UK

Dr Rawstron argued that not every patient is going to benefit from the most intensive therapy and there are safety considerations for an optimal risk–benefit ratio. He underscored that a threshold of 10^{-4} is accepted as being the minimum clinically relevant MRD. There is a wealth of data at that level allowing for prospective validation of MRD 10^{-4} as an independent prognostic marker of survival outcomes. In addition, all hospitals can be expected to measure MRD at that level.

[The International Myeloma Working Group \(IMWG\) consensus criteria for response assessment](#) defined MRD as negative below a threshold of 10^{-5} . This is the current target for FL/qPCR assays. The IMWG recommends acquiring at least 5 million cells in order to minimize the chance of a false-negative result. An LOD of 10^{-5} allows for reproducible results at 10^{-4} . However, there are some practical limitations, as multiparameter flow cytometry (MFC) is achievable only in some hospitals and the method is costly. In addition, there are limited prospective data at that level.

Assessment of MRD with a threshold of 10^{-6} provides the highest sensitivity and MRD-negative patients have the best survival outcomes. There are, however, some standardization and quantification issues. If performed well, high-sensitivity methods provide reliable information at the 10^{-4} and 10^{-5} levels and should be targeted for technical reasons. The current practice in Leeds is to assess MRD 10^{-4} to 10^{-5} on fresh samples, with DNA for high-throughput sequencing (HTS) assessment of MRD 10^{-6} .

Dr Rawstron argued that a defined MRD threshold is possibly too restrictive for comparing drugs. Usually only a small proportion of patients achieve MRD 10^{-6} negativity in clinical trials, e.g. of the daratumumab-treated patients in the [POLLUX trial](#), 29% achieved MRD 10^{-4} negativity versus 10% for MRD 10^{-6} . A [study of Myeloma IX](#) showed that there was approximately 1-year survival benefit per log depletion in residual disease. Taken together, targeting MRD 10^{-6} is the strongest prognostic marker of outcome, but may not be suitable to compare new compounds in a clinical trial. Dr Rawstron highlighted that residual disease is a biological continuum and this should be reflected in clinical trials similarly to the way clinical response rates are reported as CR, VGPR, and PR. He proposed that clinical trials report MRD4 (MRD 10^{-4} , or below 1 in 10^4 cells), MRD5 (MRD 10^{-5} , or below 1 in 10^5 cells), or MRD6 (MRD 10^{-6} , or below 1 in 10^6 cells). This approach would allow for the most comprehensive assessment of investigational drugs as well as comparing information with previous trials.

Flow cytometry

Bruno Paiva, Hematology and Immunology Departments, Clinica Universidad de Navarra, Flow Cytometry Core - CIMA LAB Diagnostics, Universidad de Navarra, EuroFlow Consortium, Spanish Myeloma Group (GEM), Spain

Dr Paiva gave an overview of the development of FC as a validated method of MRD assessment. First-generation FC had 4 colours and a sensitivity of 10^{-4} . The clinical benefit of achieving MRD negativity was applicable to all patients and subgroups, including transplant-eligible (TE) and -ineligible (TI), disease stage according to ISS, and cytogenetic risk. Noteworthy was that for patients with high-risk MM (as defined by fluorescence in situ hybridization [FISH] results), MRD negativity was the only biomarker that confirmed a favourable prognosis. With the introduction of second-generation FC (8-colour; 10^{-5}), Dr Paiva's lab was able to demonstrate that in the absence of further treatment, the prognosis for patients who are MRD 10^{-5} positive is identical to the outcome of patients with MRD levels 10^{-4} or higher. Next-generation flow cytometry (NGF; 10-colour; 10^{-6}) is a highly sensitive method for detection of MRD. It has been validated but not yet standardized. A limitation of the method is that lower MRD levels are more susceptible to variability in case of haemodilution. NGF reaches 10^{-6} LOD in 88% of patients, and 10^{-5} LOD in all patients. With this increased sensitivity, 25% of the patients defined as MRD-negative by second-generation FC were reclassified as MRD-positive, and this had an impact on prognosis.

Next-generation sequencing

Dr Hervé Avet-Loiseau, Toulouse, France

Dr Avet-Loiseau presented NGF data from the [IFM DFCI 2009 Trial](#), which included 700 TE patients with NDMM. Patients were randomized to either lenalidomide–bortezomib–dexamethasone (RVD) or high-dose therapy (HDT+ autologous transplant). All patients received 12 months of lenalidomide maintenance, and MRD 10^{-6} was assessed before and after maintenance. MRD negativity at 10^{-6} was predictive of improved PFS and OS and this was independent of treatment arm and cytogenetic high risk status. However, there were more MRD-negative patients in the transplant arm (60% versus 40% with RVD). Dr Avet-Loiseau recommended inclusion of MRD assessment with 10^{-6} as the required sensitivity

level to properly evaluate the efficacy of treatment, and with longer follow-up to see whether it translates into cure.

Discussion

- The need to assess extramedullary disease was discussed. This has never been observed in patients achieving MRD negativity at 10^{-6} , in the experience of the panel. Per the IMWG definition, patients in CR do not have any soft-tissue plasmacytomas. Regulators will likely require exclusion of extramedullary plasmacytoma (EMP) as a secondary endpoint. However, application to clinical practice would be challenging, as MRI is not part of the standard diagnostic algorithm.
- In terms of RCT evidence, the regulators would like to see more than one pivotal phase 3 trial, or if there is only a single trial, it has to be large enough (approximately 700 patients; 350 per arm) and be powered to detect significant differences/test the right hypotheses. A comment was made that when designing clinical trials that assess MRD in CR patients, the attrition rate due to BM biopsy failure and inadequate samples needs to be considered.
- It was queried whether the regulatory approach for B-cell maturation antigen (BCMA) chimeric antigen receptor (CAR) T-cell therapies would be different as this is a cellular-based approach and a potential curative option. Dr Flores answered that this would be assessed on a case-by-case basis, but she would still expect to see a (pre-defined) difference in MRD between the investigational and control arms. It is unlikely that approval would be granted based on a single-arm study even in the case of exceptional data (i.e. high CR rate in the advanced setting); randomization and a control arm would be required from the safety perspective. However, the disease setting and population are highly relevant as they set the bar for the expected level of response and durability of response in the absence of treatment. It was mentioned that Novartis has received FDA approval of a CD19 CAR T-cell therapy for ALL based on a single-arm study. While in ALL sustaining remission for 1 year is equivalent to cure, in MM it is a different scenario because we do not know whether MRD negativity sustained for a certain period would translate into cure.

CURRENTLY AVAILABLE DATA: HOW TO DEFINE THE OPTIMAL DATASET FOR VALIDATION?

Overview of 2 published meta-analyses

Walter Gregory, Clinical Trials Research Unit, University of Leeds, UK

Dr Gregory mentioned 2 recent meta-analyses evaluating the utility of MRD detection in patients with NDMM ([Landgren et al. *Bone Marrow Transplant.* 2016](#); [Munshi et al. *JAMA Oncol.* 2017](#)). For his talk, Dr Gregory focused on the Munshi et al. analysis because it was more comprehensive (15 studies included in the quantitative analysis compared with 4 in the analysis by Landgren et al). Moreover, the results from Landgren et al. were very similar in terms of reported overall hazard ratios (HR). Munshi et al. reconstructed individual patient data from the published Kaplan–Meier graphs for PFS and OS. Fourteen studies (N = 1,273) reported the impact of MRD on PFS with an overall HR of 0.41; 12 studies (N = 1,100) provided OS data according to MRD status, with an overall HR of 0.57. For patients achieving conventional CR, the overall HR was 0.44 for PFS across 5 studies (n = 574) and 0.47 for OS across 6 studies (n = 616) favouring MRD negative patients. Dr Gregory concluded that MRD is a consistent marker of long-term outcomes, including in patients with CR. He also demonstrated that by characterizing and quantifying the residual disease which regrows and causes progression following treatment, using mathematical modelling techniques, one might be able to predict the PFS curves in trials.

PETHEMA/GEM Flow-MRD datasets

Bruno Paiva

Dr Paiva gave an overview of the MRD datasets available in Spain. A total of 1,216 of the 2,132 patients (57%) had MRD assessment by FC, including 631 of 838 patients (75%) with CR (GEM2000, GEM2005MENOS65, GEM2005MAS65, GEM2010MAS65, GEM2012MENOS65 studies). The population represents a mix of TE and TI patients treated with bortezomib and/or lenalidomide containing regimens with or without maintenance. There is a particular focus on the quality of the samples and prospective well-defined points of sample collection in order to minimize missing data, which has improved over time. For example, MRD has been assessed by NGF in all patients attaining CR (n = 277; 100%) in the GEM2012MENOS65 study (bortezomib–lenalidomide–dexamethasone [VRD] induction followed by HDT plus ASCT and VRD consolidation). The trial assessed MRD at multiple time points after induction,

after HDT, and after consolidation; data will be presented at the American Society of Hematology (ASH) 2017 meeting. The dataset represents a “work in progress” and is fully available for any initiative that aims to establish MRD as a surrogate. However, the majority of the studies were phase 2, and there was no comparison of the investigational agent versus control to allow an evaluation of how differences in MRD translate to PFS differences.

Myeloma MRD: UK data

Roger Owen, St James's Institute of Oncology, Leeds, UK

The UK MRD data is a rich dataset covering a range of clinical scenarios. These include various NDMM settings: TE, TI, maintenance lenalidomide/thalidomide, high-risk, deferred autograft (Myeloma IX, Myeloma XI/XI+, PADIMAC, CARDAMON, MUK IX); and RRMM: salvage ASCT, etc. (Myeloma X, MUK five, ACCORD/Myeloma XII). Dr Owen showed the prognostic impact of achieving MFC-MRD negativity on PFS and OS using Myeloma IX data (N = 397) ([Rawstron et al. J Clin Oncol 2013](#)). There was also significant improvement in OS benefit with each log depletion of disease. Log MRD reduction as a continuous variable and cytogenetic risk profile emerged as the only 2 independent variables that were predictors of OS in multivariate analysis.

Myeloma XI/XI+ (N = 2,998) has a wealth of MRD data of TE and TI patients assessed at different time points: post-induction (n = 1,500), post-ASCT (n = 900), and after 6 months of lenalidomide maintenance (n = 500). Approximately 15% of patients in the TI subset receiving either thalidomide- or lenalidomide-based induction (N = 297) achieved 10^{-4} MRD negativity. The PFS for these patients almost doubled (HR 0.44; $p < 0.0001$). This effect was independent of the induction received.

Maintenance setting data from Myeloma XI (N = 397) will be presented at ASH 2017, showing 80% risk reduction of progression or death for patients who were MRD-negative after 6 months of lenalidomide maintenance. The very low HR suggests that persistence of MRD negativity through the maintenance period is a very important prognostic factor.

Patients in the Myeloma X trial were randomized to a second transplant at relapse or cyclophosphamide consolidation (N = 184 patients were randomized; 90 patients had MRD assessment at day 100). Patients achieving MRD negativity had a 61% risk reduction for progression or death (HR 0.39). In the MUK five trial, 292 patients with primary refractory MM were randomized to bortezomib or a carfilzomib-based reinduction therapy. MRD was assessed in 182 patients (approximately 62%), and 18% (approximately 10% IIT) were 10^{-4}

MRD negative. The relatively low number of patients with 10^{-4} MRD negativity suggests that lower thresholds (e.g. 10^{-6}) of sensitivity may not be able to demonstrate benefit in the relapsed setting.

**In-depth look at specific datasets; which ones meet requirements? Individual data
“owners”/cooperative groups**

Stefania Oliva, Myeloma Unit, Division of Hematology, University of Torino, Italy

Dr Oliva presented MFC MRD data in TE patients with NDMM treated in the [EMN02/HO95 phase 3 trial](#). In this trial, patients were randomized to either HDT plus ASCT or bortezomib–melphalan–prednisone (VMP) intensification followed by a second randomization to VRD consolidation or no consolidation. All patients received lenalidomide maintenance until progression. The 10^{-5} MRD was assessed before the start of maintenance therapy (after intensification and/or after consolidation), and then monitored every 6 months until clinical relapse. Of 316 patients in CR (stringent complete remission [sCR], CR, or nCR), 76% achieved MRD negativity. MRD-positive patients at pre-maintenance had a second MRD evaluation after 1 year on lenalidomide maintenance: 44% achieved MRD negativity after 6 to 12 months, and 4% after 18 to 24 months. Patients achieving MRD negativity had better PFS (HR 0.33) regardless of treatment arm. The findings were consistent across high-risk patients and advanced ISS disease stages, as previously discussed by Dr Paiva. Landmark analysis at 1 year post-maintenance showed a PFS benefit for MRD-negative patients versus MRD-positive patients, and at 24 months, 92% versus 65% were progression-free. Also, 85% of these patients were persistently MRD-negative (from before maintenance), suggesting the importance of sustaining MRD. The study confirmed overall that: 1) MRD by MFC is a strong prognostic factor in MM patients receiving intensification with novel agents or transplant; 2) the achievement of MRD negativity suggests the importance of deep response in high-risk MM patients defined by FISH abnormalities and ISS stage.

Comment: The false-positive rate of conversion to MRD negativity post-intensification in the absence of maintenance seems to be approximately 3%, according to the Leeds data.

Discussion

- Is there a need to use 2 different validated assays from regulatory perspective?
Currently, this is a recommendation, not a requirement. The Spanish and French

MM groups are performing NGS and NGF on the same samples according to their latest trial protocols.

- The regulators will require data from all target patient populations before they accept MRD as a proper validated surrogate endpoint.
- There was a question to the regulators regarding the recommended minimal follow-up time for MRD. The minimum follow-up for MRD would be the same as the one required to obtain PFS data. Also, the volume of data is relevant.
- There are also MRD-positive patients who have a long PFS, these patients are potentially reverting to a monoclonal gammopathy of undetermined significance (MGUS) state. There is limited data about the underlying biology of this low-risk disease. Studies aiming to understand the immune signature of these cases are ongoing.
- The microenvironment also plays a significant role and therefore it is a challenge to establish a single surrogate marker.
- What is needed to establish such a persistence of MRD negativity as a robust predictor of survival?
 - From regulatory perspective, it would be important to conduct assessments at multiple time points in order to understand the full picture. Sustained MRD will likely be as important as sustained CR (the IMWG now requires a second confirmation to define a patient in CR).
 - From a biological standpoint, sustained MRD probably means that the actual log kill on the disease was down from 10^{-6} to 10^{-8} and it takes a long time for it to regrow back. Outcomes seem to be best for those patients who achieve negativity early. In patients who do not achieve as great a level of disease reduction with the first component of the therapy, even though they get a further residual disease reduction with treatment, it does not get them down to the same level as the better-responding patients.
 - Defining the optimal time point of MRD assessment is still a matter of debate; however, Dr Paiva suggested that an informed estimate would be the time at which most patients can have an MRD assessment before progressions start to occur.

CLINICAL DEVELOPMENT STRATEGY FOR VALIDATION OF MRD AS A SURROGATE ENDPOINT

Quantifying surrogacy

Walter Gregory

The goal of surrogacy validation is to establish the relationship (coefficient of determination, R^2) between the surrogate (MRD) and the true endpoint (PFS) in terms of treatment effects. R^2 indicates the strength of relationship between the 2 outcome measures, i.e. the ability of the treatment effect on the surrogate to predict the treatment effect on the true outcome measure. $R^2 = 1$ implies perfect surrogacy that can replace the true outcome. Prof. Gregory underscored that the goal is not to replace PFS as the primary outcome, but to get a reliable intermediate endpoint to suggest further treatments or treatment alterations (e.g. switching of treatments, early stopping, etc). However, it is not clear yet what value of R^2 should be assumed.

Surrogacy is assessed at the trial level and the individual patient level: R^2 for the trial, and R^2 for the individual using meta-analytic individual participant data. The true and surrogate outcomes are paired data, i.e. from the same person. Looking at the joint distribution of the data is important because it avoids ecological bias – e.g. when a treatment effect is observed for both the surrogate and the true endpoint, but the patients who have treatment benefit on the surrogate are *not* those who have treatment benefit on the true endpoint. The joint modelling is not straightforward statistically and it depends on the type of data: continuous versus categorical data, e.g. time-to-event endpoints. At the trial level, it is also important to get a sizeable number of datasets (at least 5, ideally 15 or more; sizeable subgroups could be seen as separate units). For example, the Myeloma XI dataset can be split in 10 different data units based on the 2 treatment pathways and according to the randomizations in the multifactorial trial design.

Dr Gregory ended his talk with a summary of the prerequisites for the validation of MRD as a surrogate for PFS. These include a centralized database, multiple trial datasets in the same population, individual patient data (e.g. patient characteristics, treatment received), standardized endpoint data (MRD, PFS, OS), as well a pre-agreed statistical analysis plan with full definition of endpoints. He also encouraged other investigators to share their datasets in a joint effort.

Discussion

- Dr Gormley mentioned that the FDA is encouraging sponsors to collect MRD data in their trials as a secondary endpoint. When selecting trials, or deciding to break down trials into subgroups, etc. for the purpose of surrogacy validation, it is important to avoid introducing bias, e.g. by not including homogeneous subgroups in terms of disease stage and cytogenetic risk, etc. Early engagement with the regulators is encouraged. Although this is a formal process with an explicit meeting package, the FDA is open to preliminary, conceptual discussions to ensure alignment from the trial programme start and appropriate prioritization.
- There is not sufficient information to consider sustained MRD negativity as a surrogate endpoint in high-risk populations at present. Noteworthy is that data are usually reviewed as a package by the regulatory agencies, e.g. even if the endpoint is OS, response rate and response duration are all considered.
- Dr Gormley mentioned that the lingering questions could be answered by meta-analyses, e.g. surrogate threshold, what time point, etc. The only caveat is that information from retrospective datasets cannot be extrapolated to drugs with a biologically different mechanism of action (e.g. a CR 30-month endpoint in FL is not relevant to treatment with ibrutinib, as the depth of response does not correlate with survival outcomes). Dr Flores mentioned that a prospective evaluation would be the strongest body of evidence for surrogacy.
- Dr Gormley confirmed that ORR, as a surrogate marker for PFS, is currently used for accelerated approvals, provided that it is of sufficient magnitude and adequate durability with confirmatory trials, or continuing to file for approval on the initial trial for PFS.
- The need for novel efficacy and safety biomarkers to select patient populations likely to respond to treatment was discussed as an area of extreme importance.

MRD AS A SURROGATE ENDPOINT: INSIGHTS FROM CLL

Andy Rawstron

Ensure broad access to techniques for measuring MRD and identifying suitable approaches to reporting response

Dr Rawstron discussed the approval of MRD as a regulatory point in CLL and highlighted that this was a long process that took approximately 20 years, accelerated by the introduction of regimens that can induce MRD-negative CRs, such as venetoclax.

MRD after therapy is an important predictor of outcome in patients with CLL and there is a linear improvement in PFS per log depletion in MRD.

At present, undetectable MRD in patients with clinical CR after induction therapy is accepted as an intermediate endpoint for EMA licensure in RCTs designed to show superiority in terms of PFS ([Appendix 4 to the guideline on the evaluation of anticancer medicinal products in man EMA/CHMP/703715/2012 Rev. 2](#), available at www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2016/02/WC500201945.pdf).

The [2008 International Workshop on CLL guidelines](#) defined MRD 10^{-4} as the threshold; however, newer treatments are inducing deeper remissions, and there is a need to increase the sensitivity of the available assays. As one improves the method, it has to remain compatible/comparable with the available trial data for regulatory purposes.

The [European Research Initiative on CLL \(ERIC\)](#) led by Dr Rawstron validated an FC approach to reliably quantitate MRD at the level of 10^{-5} . The assay comprised a core panel of 6 markers which can be re-validated by most labs using normal PB. This method is directly comparable to previous ERIC-designed assays. Although HTS offers higher sensitivity (10^{-6}), with good concordance with FC results at the 10^{-4} level, Dr Rawstron argued that ensuring broad access to the techniques is critical to collecting a large amount of trial data. His recommendation is:

- Apply the guidelines to MRD4 ($< 10^{-4}$)
- Measure to MRD5 ($\leq 10^{-5}$)
- Explore MRD6 (10^{-6})

Understand compartment effects

The role of PB as a source for MRD analysis has also been evaluated. There is some correlation between the levels of disease in PB and BM and on average there is a log difference between the 2 compartments, i.e. a compartment effect. Although PB is acceptable for monitoring during and after therapy for CLL (short-term outcomes), BM may be required if there is a significant compartment effect, which is the case for some drugs, e.g. alemtuzumab, rituximab plus chemotherapy. BM is required for a definitive MRD response assessment as a surrogate of long-term outcomes assessment.

Improve knowledge of MRD kinetics and biological responses to optimize treatment

Individual patient responses to a given treatment regimen may differ in the extent and rate of depletion of cancer cells until a plateau is reached. Therefore, it is important to measure MRD at many different time points in order to understand these kinetics. Capturing this information in theoretical models may help us guide treatment and monitor response, determine optimal treatment duration, and delay development of resistance. For example, stopping therapy after doubling the time required to reach MRD negativity results in deeper and more extended remissions, compared with stopping treatment 6 months after confirmed MRD negativity in UK CLL trials. In addition, identifying the biology of response at an MRD level would allow a better understanding of the biology of relapse and the requirements for the next set of treatment.

SUMMING UP AND KEY CONCLUSIONS

John Smyth, Professor Emeritus of Medical Oncology, University of Edinburgh, UK

“For myeloma, MRD would be central in the definition of cure.” — Hervé Avet-Loiseau

“Some things are going to be tricky.” — Andy Rawstron

- The challenge with MRD as an endpoint is trying to quantify a very complex biological event. The vast majority of data we have collected over the course of the last 20 years is confounded by treatment advances in the field.
- MRD positivity in PB correlates with MRD assessments in BM. Those patients could be spared painful BM biopsies.

- All hospitals at the present time could be expected to assess MRD at 10^{-4} , but 10^{-5} should be the target. MRD 10^{-6} is a better goal in the research setting. The oncology working party has agreed so far on a cut-off of 10^{-5} , which seems to be acceptable to the EMA.
- Regulatory agencies are not prescribing a specific technology, but whichever is used has to be standardized and validated. The EMA recommends using 2 different assays in the same trial.
- Currently, regulators are considering MRD only in the context of CR. This might need some further discussions, as does the need to exclude extramedullary disease.
- In the USA, MRD is accepted as a prognostic tool, and in the setting of regulation, for patient stratification and possibly for patient selection and risk-based treatment assignments, but not yet MRD as a surrogate endpoint.
- MRD negativity is predictive of long-term survival outcomes with consistent hazard ratios, reducing the risk of progression or death by half. The meta-analysis of MRD data across trials had large patient numbers (among patients in CR, 306 patients were MRD-negative and 178 were MRD-positive).
- There are large datasets of MRD data in the UK, Spain, France and other countries from the European Myeloma Network.
- The role of MRD assessment with CAR T-cell therapies, where the treatment goal will likely be cure, remains to be elucidated.
- The role of MRD assessment in patients with high cytogenetic-risk MM requires more data.
- There is a room for a pre-competitive initiative between the industry, academia, and regulators to align efforts in answering some of the key questions around MRD surrogacy. There is already a consortium in place led by the Mayo Clinic aiming to gather available MRD data.
- As there are a number of challenges to retrospective approaches (i.e. inconsistency in data collection), another approach can be to address the regulatory requirements for MRD surrogacy in prospective studies. An important prerequisite for this approach is that there are good definitions in place.
- It is unclear whether HTAs would accept MRD as an endpoint.

ABBREVIATIONS

ALL, acute lymphoblastic leukaemia

AML, acute myeloid leukaemia

ASCT, autologous or allogeneic stem cell transplantation

ASH, American Society of Hematology

BCMA, B-cell maturation antigen

BM, bone marrow

CAR, chimeric antigen receptor

CDDF, Cancer Drug Development Forum

CHMP, Committee for Medicinal Products for Human Use

CLL, chronic lymphocytic leukaemia

CML, chronic myeloid leukaemia

CR, complete remission

CTC, circulating tumour cell

DFS, disease-free survival

ELDA, extreme limiting dilution analysis

EMP, extramedullary plasmacytoma

FC, flow cytometry

FDA, Food and Drug Administration

FISH, fluorescence in situ hybridization

HDT, high-dose therapy

HTS, high-throughput sequencing

IIT, intention to treat

IMWG, International Myeloma Working Group

IND, investigational new drug

ISS, International Staging System

IWG, International Working Group

LOD, limit of detection

MCL, mantle cell lymphoma

MFC, multiparameter flow cytometry

MGUS, monoclonal gammopathy of undetermined significance

MHRA, Medicines and Healthcare Products Regulatory Agency

MM, multiple myeloma

MRD, minimal residual disease

NCI, National Cancer Institute

nCR, near complete remission

NDMM, newly diagnosed multiple myeloma

NGF, next-generation flow cytometry

NGS, next-generation sequencing

ORR, overall response rate

OS, overall survival

PB, peripheral blood

PFS, progression-free survival

PR, partial response

RCT, randomized controlled trial

RR, response rate

RRMM, relapsed/refractory multiple myeloma

RVD, lenalidomide–bortezomib–dexamethasone

sCR, stringent complete remission

TE, transplant eligible

TI, transplant ineligible

VGPR, very good partial response

VMP, bortezomib–melphalan–prednisone

VRD, bortezomib–lenalidomide–dexamethasone

WM, Waldenström macroglobulinaemia