The Level of Minimal Residual Disease in the Bone Marrow of Patients with Multiple Myeloma before High-Dose Therapy and Autologous Blood Stem Cell Transplantation Is an Independent Predictive Parameter

Mark Korthals,1,3 Nina Sehnke,1 Ralf Kronenwett,1,* Ingmar Bruns,1 Jochen Mau,2 Fabian Zohren,1 Rainer Haas,1 Guido Kobbe,1 Roland Fenk1

The prognostic relevance of minimal residual disease (MRD) in patients with multiple myeloma is still an open question. We measured MRD levels in bone marrow (BM) samples of 53 patients treated with high-dose therapy (HDT) and autologous peripheral blood stem cell transplantation using real-time quantitative (RQ)-IgH-PCR with allel-specific oligonucleotide probes. We identified a prognostically relevant threshold level of 0.2% IgH/β-actin ratio in the BM before HDT. Twenty-six patients with MRD levels below this value were termed the “low-MRD group,” whereas 27 patients with levels above this threshold were allocated to the “high-MRD group.” Median event-free-survival (EFS) in the low-MRD group was significantly (P = .001) longer than in the high-MRD group with 35 versus 20 months, respectively. Overall survival (OS) within the low-MRD group was also significantly longer with 70 versus 45 months (P = .04). Using multivariate analysis, we found that the pretransplantation MRD level was an independent prognostic factor for EFS (P = .003) and OS (P = .05). Further, EFS of patients in the high-MRD could be improved (P = .005) when they achieved a low MRD level after HDT. In conclusion, measuring MRD is of prognostic relevance in patients with MM, and low MRD levels should be a goal of treatment.


KEY WORDS: Minimal residual disease, Multiple myeloma, RQ-IgH-PCR, Prognostic parameter

INTRODUCTION

High-dose chemotherapy (HDT) with autologous peripheral blood stem cell transplantation (PBSCT) has improved the survival of patients with multiple myeloma (MM) [1]. To estimate the prognosis of these patients, the new international staging system (ISS) [2] is commonly used, which relies on cutoff values for β2-microglobulin and serum albumin. Another relevant prognostic factor is the presence or absence of cytogenetic aberrations [3]. Still, the results of cytogenetic examinations, including classical banding techniques or fluorescence in situ hybridization, are not available for all patients with MM. A low proportion of plasma cells in the bone marrow (BM) before HDT is also reportedly associated with a longer event-free (EFS) and overall survival (OS) [4]. The same is true with regard to the achievement of a complete remission, which is defined by the absence of detectable paraprotein and a negative immunofixation in the serum and urine [5]. Both parameters are readily available with the limitation of an analytical sensitivity of approximately 10^-2.

Because drugs such as thalidomide [6], bortezomib [7], or lenalidomide [8] are quite effective and associated with a considerably higher proportion of patients with very good partial remission (vgPR) or complete remission (CR), the detection of minimal residual disease (MRD) in patients with MM could be of particular relevance. Recently, multiparameter flow cytometric
detection of aberrant plasma cells in the BM before and after HDT and autologous PBSCT was a reliable predictive factor for EFS and OS [9]. Using a limiting dilution assay for IgH-PCR, Bakkus et al. [10] identified a threshold level of 0.015% clonotypic cells in BM samples obtained 3 months after HDT and autologous PBSCT as prognostically relevant for the EFS. Others have shown that there is a 0.01% threshold level for EFS by the simultaneous measurement of MRD in patient samples with quantitative PCR and flow cytometry [11]. Further, the prognostic value of monitoring MRD by using IgH-PCR could be demonstrated in the setting of allogeneic transplantation [12] and after HDT followed by a consolidation therapy including bortezomib, thalidomide, and dexamethasone [13].

Still, the detection of MRD is no standard method to estimate the risk of relapse in patients with MM. Therefore, further studies are needed before results of MRD may permit therapeutic decisions, as shown for patients with acute lymphoblastic or chronic myeloid leukemia [14].

It is not clear whether the achievement of a molecular remission before HDT is associated with a better outcome after transplantation. This is of particular interest with regard to the introduction of new drugs into the induction therapy before HDT and autologous PBSCT leading to higher response rates [15]. To address this question, we used real-time quantitative (RQ)-IgH-PCR to evaluate the predictive power of molecular MRD monitoring before HDT and autologous PBSCT. Further, we investigated whether a molecular defined remission correlates with the conventional remission status of patients.

MATERIALS AND METHODS

Patients and Samples

The 53 patients included in this analysis suffered from symptomatic MM and received the same first-line therapy at our hospital. Between November 2000 and October 2007, we treated 121 patients with stage II/III symptomatic MM with induction chemotherapy followed by granulocyte-colony stimulating factor (G-CSF) supported cytotoxic mobilization and a single HDT with autologous PBSCT. Looking at the entire group, 80 patients presented with a myeloma subtype of IgG or IgA qualifying for MRD monitoring using IgH-PCR. BM samples could be obtained from 70 patients, and the specific IgH sequence was identified for 55 patients (78%). All patients with a suitable IgH marker were included in the study regardless of a negative or positive immunofixation or the remission status after HDT. A corresponding IgH-PCR fulfilling the below defined quality criteria could be established in 53 patients (76%). In 2 patients, the PCR did not reach the required sensitivity of at least $10^{-9}$.

The induction treatment of all 53 patients included in this analysis consisted of 2 to 4 cycles of idarubicine (20 mg/m² i.v., days 1-4) and dexamethasone (40 mg oral, days 1-4, 9-12, 17-21) in monthly intervals. The patients then received Pegfilgrastim (Amgen, Munich) or G-CSF in combination with cyclophosphamide (2 g/m², days 1+2) for PBSC mobilization. Thereafter, a single HDT was administered at a dose of 200 mg/m² melphalan followed by autologous PBSCT. Approximately 3 months after HDT, all patients received maintenance therapy with interferon alpha (n = 27) or thalidomide (n = 18). In the case of an HLA-identical sibling donor, an allogeneic transplantation was performed (n = 8) following reduced-intensity conditioning (RIC) with fludarabine 30 mg/m² on days −3 to −1 and 2 Gy total body irradiation (TBI) on day 0. Interferon alpha was given at a dose between 1.0 and $4.5 \times 10^6$ U subcutaneously twice per week. The daily dose of thalidomide varied between 100 and 400 mg. Maintenance therapy with interferon or thalidomide was continued until disease progression or discontinued in case of toxicity. Remission status and disease progression were defined according to the International Myeloma Working Group Uniform Response Criteria (IMGW-URC) [16], with the exception that a complete remission was not confirmed by BM biopsy. Because the majority of patients were treated before 2003, data on the concentrations of free-light chains in the serum were not available. Cytogenetic testing was done by conventional banding metaphase analysis. The median EFS and OS for the 53 patients were 28 and 61 months, respectively. For patients treated with interferon (IFN) maintenance, thalidomide maintenance, or RIC allogeneic transplantation, the median EFS was 20, 45, and 25 months, respectively ($P = .04$) and the median OS was 53, 66, and 40 months, respectively ($P = .6$). The characteristics of the 53 patients are summarized in Table 1.

BM samples were obtained at first diagnosis, 2 to 10 days before and 3 to 6 months after HDT and autologous PBSCT. Fifteen patients did not consent to a third BM puncture after HDT and, therefore, data of MRD after HDT is only provided for 38 patients. All patients gave written consent for treatment and molecular analysis. The study was performed according to the guidelines of the ethical committee of the University of Duesseldorf.

Sequencing of the Patient-Specific IgH Sequence and Real-Time Quantitative IgH-PCR

IgH sequences were derived from total RNA preparations of samples at the time of diagnosis and a patient-specific RQ-IgH-PCR was performed on the LightCycler system (Roche, Mannheim, Germany) using Taqman-technology as previously described...
**Table 1. Patients’ Characteristics**

<table>
<thead>
<tr>
<th></th>
<th>MRD before Transplantation</th>
<th>MRD after Transplantation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All Patients</td>
<td>&lt;0.2% (n = 26)</td>
</tr>
<tr>
<td>Age (median years; range)</td>
<td>54 (31.75)</td>
<td>54 (32.69)</td>
</tr>
<tr>
<td>Sex (m:f; %)</td>
<td>66:34</td>
<td>58:42</td>
</tr>
<tr>
<td>Ig subtype (IgG: IgA; %)</td>
<td>77:23</td>
<td>77:23</td>
</tr>
<tr>
<td>Stage D+S (I + II: III; %)</td>
<td>11:89</td>
<td>12:88</td>
</tr>
<tr>
<td>Stage ISS (1:2:3; %)</td>
<td>53:34:13</td>
<td>54:35:12</td>
</tr>
<tr>
<td>Cyto genetics (abnormal:normal; %)</td>
<td>26:74</td>
<td>23:77</td>
</tr>
<tr>
<td>PC BM infiltration before HDT (&lt;5%; %)</td>
<td>34:66</td>
<td>44:56</td>
</tr>
<tr>
<td>PC BM infiltration after HDT (&lt;5%; %)</td>
<td>50:50</td>
<td>56:44</td>
</tr>
<tr>
<td>Salvage therapy with Thal (yes:no; %)</td>
<td>58:42</td>
<td>61:39</td>
</tr>
<tr>
<td>Salvage therapy with Bor (yes:no; %)</td>
<td>49:51</td>
<td>48:52</td>
</tr>
<tr>
<td>Salvage therapy with Lena (yes:no; %)</td>
<td>32:68</td>
<td>30:70</td>
</tr>
</tbody>
</table>

BM indicates bone marrow; Bor, bortezomib; f, female; IFN, interferon; HDT, high-dose chemotherapy; ISS, international staging system; m, male; MRD, minimal residual disease; MRSD, minimal response/stable disease; nCR, near complete remission; PC, plasma cell; PR, partial remission; RIC, reduced-intensity conditioning allogeneic transplantation; D+S, Durie-Salmon; Thal, thalidomide; Bor, Bortezomib; Lena, lenalidomide.

*C*Correlations were calculated by the Fisher exact test for 2 categories or by the chi-square test for 3 categories or continuous variables.

**Determination of a Prognostic Cutoff Value**

To preserve sufficient power, the data from 10 patients of our former analysis [17] and another 43 patients observed subsequently were pooled. Determination of prognostic cutoff points on quantitative MRD was done in 3 steps: (1) compare the empirical cumulative distribution functions (ECDF) as cumulative sum percentage curves over the axis of quantitative MRD values between a positive and a negative outcome for a statistically significant (P < .05) difference with a Kolmogorov-Smirnov test; (2) determine the value of quantitative MRD with maximal distance between the 2 ECDF in any binary endpoint of a significant Kolmogorov-Smirnov test in step (1); and (3) determine the cutoff point with a receiver-operating characteristic (ROC) curve as the value of quantitative MRD with a maximal Youden index. As a result, a significant prognostic threshold level of the 2IgH/β-actin ratio from BM samples before HDT could be defined. The use of a cutoff value of 0.2% 2IgH/β-actin, as determined by ROC analysis, resulted in the most sensitive and specific prognostic estimate (Supplemental information).

**Statistical Analysis**

OS and EFS of patients were calculated from the beginning of treatment after first diagnosis and analyzed by Kaplan-Meier plots and the log-rank test. Correlations between patient characteristics were analyzed using the Fisher exact test for nominal variables with 2 categories and the 2-sided chi-square test for nominal variables with 3 categories of continuous variables. Uni- and multivariate analyses were performed using the Cox regression analysis. All patient characteristics with P ≤ .1 in the univariate analysis and a minimum of 14 patients per group entered the multivariate analysis.

**RESULTS**

**MRD Levels Do Not Correlate with Remission Status**

The PCR results from the 53 BM samples collected 1 week before HDT showed a median MRD ratio of 0.2% 2IgH/β-actin. Nine of the 53 patients had a negative PCR result, while for the remaining 44 patients the 2IgH/β-actin ratio varied between...
0.001% and 86%. To analyze the MRD levels after HDT and autologous PBSCT, samples from BM of 38 of the 53 patients were collected 3 to 6 months after transplantation. The respective median MRD ratio was 0.3% 2IgH/ß-actin. PCR results were negative for 8 patients, 3 of them with a negative result before HDT, whereas 30 patients had posttransplantation MRD values between 0.001% and 21%.

To test whether the MRD level in BM as measured by RQ-IgH-PCR is a mere molecular correlate of the IMWG-URC remission status, patients with a different remission status were compared concerning the 2IgH/ß-actin ratio in BM (Figure 1). There was no correlation of the MRD level before HDT and autologous PBSCT and the remission status before transplantation. Similarly, the MRD level after HDT and autologous PBSCT did not significantly correlate with the remission status after transplantation. Therefore, MRD assessment on a molecular level is a specific measure of residual clonal tumor cells and differs from the standard evaluation of the remission status.

### The MRD Level in Bone Marrow before High-Dose Chemotherapy Is a Prognostic Parameter for EFS and OS

We looked for a significant prognostic threshold level of the 2IgH/ß-actin ratio from BM samples before HDT. Using ROC analysis with a maximal Youden index, a cutoff value of 0.2% 2IgH/ß-actin resulted in the most sensitive and specific prognostic estimate.

In the next step, we analyzed the EFS and OS of the patients after allocation to a low-MRD or high-MRD group, according to the defined cutoff ratio of 0.2% 2IgH/ß-actin in the BM before HDT. Of these, 26 were considered to belong to the low-MRD group, whereas 27 patients were allocated to the high-MRD group. There were no significant differences between these 2 groups concerning age, sex, MM subtype, stage according to Salmon & Durie, ISS stage, cytogenetics, kind of maintenance therapy, plasma cell infiltration in the BM, and remission status before or after HDT or the kind of salvage therapy (Table 1).

Analysis by Kaplan-Meier plots and the log-rank test showed a difference in the EFS between both groups of patients. The EFS of patients with MRD ratios of <0.2% 2IgH/ß-actin was significantly (P = .001) longer than that of patients with MRD ratios of >0.2% 2IgH/ß-actin with a median of 35 months and 20 months, respectively (Figure 2A). In the same line, a difference in the OS between both groups was found. After a median follow-up of 61 months, patients belonging to the low-MRD group had a significantly (P = .04) longer OS with a median of 70 months compared with the patients of the high-MRD group who had a median OS of 45 months (Figure 2B).

### The MRD Level in Bone Marrow after High-Dose Chemotherapy Is a Prognostic Parameter for EFS

Grouping the 38 patients with samples obtained after HDT and autologous PBSCT on the basis of an MRD ratio of 0.2% 2IgH/ß-actin after HDT resulted in a significantly different median EFS with 32 versus 20 months, respectively (P = .04, Figure 2C). In contrast, the OS was not significantly different
between patients allocated on the basis of the post-transplantation MRD value ($P = .4$, *Figure 2D*). As shown in *Table 1*, there were no significant differences in the 2 groups concerning patients’ characteristics.

To test whether the posttransplantation MRD level adds further prognostic information to the pretransplantation MRD level, we analyzed the EFS among those patients who had high MRD levels before HDT after allocation with regard to the posttransplantation MRD level (*Figure 2E*). Those patients whose MRD levels were reduced to $<0.2\% \text{2IgH}/\beta$-actin ratio after transplantation had a significantly longer EFS ($P = .005$) compared with patients who still had MRD levels above this threshold after transplantation, with 41 versus 14 months, respectively. Still, the OS of these patients with low MRD levels after transplantation was not significantly prolonged (median not reached vs 34 months, $P = .4$, *Figure 2F*).

**Uni- and Multivariate Analysis of Predictive Parameters**

Univariate analysis (*Table 2*) showed that the molecular cutoff level in BM before ($P = .002$) and after ($P = .04$) HDT and autologous PBSCT as well as the kind of maintenance therapy ($P = .02$) were
predictive parameters for EFS. By multivariate analysis (Table 2), these factors were also independent factors for EFS ($P = .003$, $P = .007$ and $P = .003$, respectively). Metaphase cytogenetic analysis, which showed a borderline significance as a prognostic parameter for EFS in the univariate analysis ($P = .1$), was also an independent factor by multivariate analysis ($P = .04$).

With regard to OS, the significant predictive variables were MRD level in BM before HDT ($P = .04$) and metaphase cytogenetic results ($P = .001$). Both factors were independent factors for OS by multivariate analysis ($P = .05$, $P = .003$, respectively).

The cutoff level of 0.2% 2IgH/2-actin in BM before HDT and autologous PBSCT and metaphase cytogenetic results were the only independent factors predictive for both EFS and OS. The 2IgH/2-actin threshold ratio therefore adds prognostic information to established parameters such as ISS stage and cytogenetics. This is true not only for patients with good prognostic parameters, such as low ISS stage and normal karyotype, but also for patients with poor prognosis as defined by a high ISS stage or an abnormal karyotype (Figure 3). Remarkably, the poor prognostic influence of an abnormal karyotype confirmed by conventional banding analysis was not valid any longer in 5 patients who achieved a low MRD status before HDT.

**DISCUSSION**

The results of our study demonstrate the predictive value of molecular MRD monitoring in patients with MM. On the basis of 53 patients, we could define a cutoff ratio of 0.2% 2IgH/2-actin in BM before HDT and PBSCT as a relevant predictive threshold for the probability of EFS. Patients falling below this threshold at the time before HDT also had a better OS than patients with values above this cutoff level. The MRD level was prognostically independent of other established prognostic parameters such as ISS stage or metaphase cytogenetic results. As a consequence, our data imply that a low pretransplantation MRD level should be a goal for induction treatment.

Our finding that a low MRD level before HDT and autologous PBSCT is associated with a prolonged EFS and that patients in the “poor prognosis” group
with high MRD levels can be further divided in 2 prognostic groups according to their MRD status after HDT, are in line with the results of a recent study using multiparameter flow cytometry for the detection of MRD [9]. In a subgroup analysis by Paiva et al. [9] 157 patients were evaluated in 3 groups. Patients who were persistently MRD positive had the worst prognosis. Patients who were MRD positive before PBSCT but improved to MRD-negative after PBSCT had an intermediate prognosis, and patients who achieved optimal response early and who were MRD negative before transplantation had the best prognosis. The MRD threshold level in this study and another study was 0.01% or lower as determined by the detection limit of the flow cytometry method [9,11]. The difference between this cutoff level and the threshold level of 0.2% in our study is most likely related to the different techniques of enumerating clonotypic cells. Keeping in mind that changes of MRD levels are assessed in log steps, the threshold level in our study are comparable to other published data [9-11]. Altogether, our data confirm the prognostic value of MRD monitoring and encourage the use of MRD measurement in patients with MM in the future.

Figure 3. The pretransplantation MRD cutoff level of 0.2% 2lgH/β-actin ratio is an independent prognostic parameter and provides additional prognostic information to ISS stage and cytogenetics. EFS of patients with high or low MRD levels was separately analyzed for patients with either normal (A) or abnormal (B) karyotype by Kaplan-Meier plots and log-rank test. Similarly, EFS of patients within the high and low MRD group was separately analyzed for patients with either ISS stage 1 (C) or patients with ISS stages 2 or 3 (D).

From a clinical point of view, MRD results not only provide a new independent parameter to assess the prognosis of patients, but also help us to direct therapeutic decisions in the future. Our findings show that a molecular remission with a pretransplantation MRD level below 0.2% 2lgH/β-actin is desirable before the patient receives a HDT. A more intensive induction therapy including thalidomide, bortezomib, or lenalidomide could be helpful to achieve this aim [8]. Notably, the data presented here derive from a treatment strategy with conventional chemotherapy, and therefore a low MRD level may only be a surrogate marker for chemosensitive disease and an effective induction therapy. In the future, it will be interesting to compare our MRD data from an earlier era of conventional induction chemotherapy with MRD data of patients receiving novel agents as initial therapy. Awaiting further evaluation is whether a low level of MRD following induction therapy with these agents is still of predictive value. Comparing the MRD level before and after transplantation favored the MRD level before HDT as a more significant predictive factor for OS in our analysis. Notably, this discrepancy was not related to differences in population selection concerning rescue strategies or other patient characteristics.
However, keeping in mind that less patient samples were available for the evaluation of the MRD level after HDT and that deaths occur more rarely than disease progressions, this comparison lacks the necessary statistical power. Our results of MRD on a molecular level are perfectly in line with clinical data showing that the achievement of a vgPR to induction treatment was a better prognostic factor in the IFM 2005-01 trial than vgPR as the best response after HDT [18].

It was interesting to note that the level of clonotypic cells was not related to the remission status as defined by serum electrophoresis, immunofixation, and the proportion of plasma cells in the BM before HDT. This is in line with other studies showing no relationship between the remission status or the number of plasma cells in the BM and the treatment outcome of patients [19-21]. Moreover, this finding is consistent with a recently published comparison of the remission status as defined by immunofixation, serum-free light chains, and immunophenotyping. In this study with 102 patients, the results gained by these different methods did not correlate, indicating that these methods provide complementary data [22]. One explanation for the lack of correlation between a negative MRD determined on a molecular level and a detectable M-protein in the serum, or even a detectable plasma cell infiltration of the BM, is that the lifespan of an individual paraprotein varies from patient to patient, with a range between a few days and 3 weeks [23]. As a consequence, the paraprotein is eventually still detectable, whereas the malignant cell clone is significantly reduced. Moreover, there are difficulties in discriminating normal or reactive plasma cells from the malignant cell clone by BM aspiration. Because multiparameter flow cytometric analysis can accurately distinguish between normal and aberrant plasma cell, the prognostic impact of this technique is, similar to the PCR-based MRD assessment, superior to the classical evaluations of remission status [9]. On the other hand, the PCR-based detection of MRD without overt signs of active disease as assessed by serologic or histologic examination is probably related to clonal cells without the morphology of a plasma cell [24-27]. The emergence of such cells may be a result of changes in the differentiation state or genetic exchanges between neoplastic and microenvironmental cells, as already discussed by Ladetto et al. [13]. Therefore, the molecular assessment of MRD is not only a very sensitive but also the most specific method in this context, and thus may add further information to flow cytometric MRD assessment. A comparison of MRD results gained by flow and RQ-PCR on samples of 24 patients showed that both techniques provide similar prognostic information [11]. As both methods for MRD assessment detect different target cells, a comparison on a larger number of patient samples is needed. However, when considering the broader availability and limited costs [11], it is more likely that MRD assessed by flow cytometry will be the method of choice for routine diagnostics in the future. Further, it should be considered that only 53 of our 122 patients had an informative PCR as a result of the need for a BM sample at the time of diagnosis and the restriction to patients with a heavy-chain myeloma subtype. Therefore, molecular MRD detection will be reserved for special scientific questions in the future.

In conclusion, MRD detection by RQ-IgH-PCR is of prognostic relevance and provides a rationale and basis for a patient-tailored therapy dependent on the individual response to a given treatment.

ACKNOWLEDGMENTS

This work was supported by Deutsche Krebshilfe (Bonn, Germany), Leukämie Liga e.V. (Duesseldorf, Germany), and Novartis.

AUTHORSHIP STATEMENT

R.F. designed research, collected data, performed statistical analysis, interpreted data, and drafted the manuscript. M.K. performed the experiments and statistical analysis, interpreted data, and drafted the manuscript. N.S. performed the experiments and collected data. R.K. designed research, contributed substantially to the development of the methodology and interpreted data. I.B. was involved in patient recruitment and correlation of clinical data with experimental findings. J.M. performed statistical analysis. F.Z. was significantly involved in patient recruitment and substantially contributed to the development of the methodology. R.H. was significantly involved in the induction and conception of the study and interpreted the data. G.K. was significantly involved in patient recruitment and correlation of clinical data with experimental findings. All authors critically revised the manuscript.

Financial disclosure: G.K. was consultant to Amgen and Celgene, received lecture fees by Amgen, Celgene, and Janssen-Cilag, and received research funds from Amgen and Celgene. R.F. was a consultant to Celgene, received lecture fees by Celgene and Janssen-Cilag, and received research funds from Novartis and Celgene.

REFERENCES


SUPPLEMENTARY METHODS

Amplification and Sequencing of the Patient-Specific IgH Sequence

Mononuclear cells (MNCs) were purified by standard Ficoll density gradient centrifugation. Genomic DNA was isolated from BM samples using the QIAamp Blood Kit (Qiagen, Hilden, Germany), and total RNA was extracted using the RNeasy Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer’s instructions. IgH sequences were derived from total RNA preparations of samples at the time of diagnosis. For amplification of the variable region of the IgH sequence, total RNA was reversely transcribed, and a consensus PCR was performed with an FR1c IgH sequence, using the Gene Amp RNA PCR core kit (Applied Biosystems, New Jersey). PCR products of all patients were cloned by ligation to the pCR4-TOPO plasmid vector and transformation of E.coli bacteria using the TOPO T/A Cloning Kit for Sequencing (Invitrogen, Paisley) according to the manufacturer's recommendations. Plasmid DNA from 10 bacterial clones of each patient was purified using the QIAprep Miniprep Kit (Qiagen, Hilden) and sequenced by a commercial supplier (SEQLAB, Göttingen). A sequence repeated in at least three different clones was considered to be the patient-specific IgH sequence. The patient-specific complementarity determining regions (CDR) 1 to 3 were identified by comparing the IgH sequence to published VH, D and JH regions (CDR) 1 to 3. An appropriate consensus primer was chosen to a consensus primer complementary to the CDR3 region, an ASO sense primer complementary to the CDR2 region, and an ASO antisense primer (40% of patients) complementary to the CDR3 region were designed with the help of OLIGO 6.0 software or BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi) by and purchased from TIB Molbiol (Berlin, Germany). ASO primers and Taqman probes were complementary to sequences with a maximum frequency of patient specific germline IgH mutations and minimal G:C content. For 60% of patients who had a short CDR3 sequence, an appropriate consensus primer was chosen as antisense primer.

The IgH-PCR was performed according to the ESG-MRD-ALL guidelines (3) with some exceptions as stated below. The reaction was carried out in a final volume of 20 μl containing 1xLightCycler-FastStart Hybridisation Probes Reaction Mix (Roche, Mannheim), optimized MgCl2 of 4 to 8 mM dependent on the target sequence, 0.5 μM of the corresponding sense and antisense primers, 200 nM ASO Taqman probe and 1μg of genomic DNA isolated from BM samples. Cycling conditions were as follows: one cycle of pre-incubation at 95°C for 10 min, 45 cycles of a denaturation step at 95°C for 15 sec and an amplification step at 60-67°C for 30 sec, one cycle of cooling at 40°C for 5 min. LightCycler software was set as follows: acquisition mode at the end of the amplification: “single”, channels: “F1/F2”, fluorimeter gains: “automated”. Each run contained a non-template control. Instead of serial dilutions of DNA from the diagnostic samples with known plasma cell infiltration as recommended by the ESG-MRD-ALL guidelines, a serial 10-fold dilution of 105 to 103 plasmid IgH copy numbers was used as an external, exogeneous standard. This modification was done due to our previous observation (2) that the molecular MRD level did not perfectly correlate with the proportion of plasma cells within BM samples. For amplification, linearized plasmid DNA was used in the presence of 660ng genomic DNA pooled from samples of healthy donors, corresponding to approximately 105 cells. The analytical sensitivity of the reaction was tested for all patients with a serial 10-fold dilution of the plasmid standard ranging from theoretically 106 to 105 plasmid copy numbers. Each patient-specific PCR reaction required a minimal sensitivity of 10-4 (10-5 in 88% of cases) and a linear amplification within the quantitative range with r ≥ 0.98. For all patients, the PCR reaction had a linear analytical sensitivity of ≤ 10-4 corresponding to a quantitative range from 10 clonal cells out of 105 total cells. To confirm analytical specificity, background amplification of each IgH-PCR was tested with a pool of DNA from healthy donors. Using the cycling conditions described above, no background amplification could be detected for each IgH-PCR. Positive PCR results were defined by at least one specific amplification of two replicates, as determined by the shape of the amplification curve and a Ct value set apart from the background level. Negativity was assured by PCR without specific amplification in 5 replicates.

As a reference for relative quantification, a β-actin PCR was run on the same instrument with the same DNA samples using the Taqman β-Actin Control Reagents Kit (Applied Biosystems, Foster City) according to the manufacturer's instructions. Cycling conditions were identical to those described for the IgH-PCR.

The IgH and β-actin copy numbers within a sample were calculated from the mean Ct values by the use of the respective standard curve for the IgH- or β-actin PCR, respectively. Mean values of the copy numbers
from two replicates were used and the proportion of clonotypic cells among MNCs was expressed as the 2IgH/ß-actin ratio in percent. PCR positive results that were outside the quantitative range and thus not quantifiable were set to the theoretical minimum 2IgH/ß-actin ratio of 0.001% (1 clonal cell out of 10^5 total cells). The cut-off level of 0.2% 2IgH/ß-actin ratio was within the quantitative range of each IgH-PCR.

**Determination of a Prognostic Cut-Off Value**

Event-free survival and overall survival times were defined as the time from beginning of treatment after first diagnosis to the date of disease progression or death, respectively. To preserve sufficient power, the data from 10 patients of our former analysis (2) and another 43 patients observed subsequently were pooled. This does not need any formal statistical adjustment as the decision to treat more patients did not depend on the analysis of the first 10 patients.

To estimate the optimal threshold value (cut-off point) of a diagnostic marker, the Youden index is commonly used. Therefore, determination of prognostic cut-off points on quantitative MRD was done in three steps: (i) compare the empirical cumulative distribution functions (ECDF) as cumulative sum percentage curves over the axis of quantitative MRD values between a positive and a negative outcome for statistically significant (P < 0.05) difference with a Kolmogorov-Smirnov test, (ii) determine that value of quantitative MRD with maximal distance between the two ECDF in any binary endpoint of a significant Kolmogorov-Smirnov test in step (i), and (iii) determine the cut-off point with a receiver-operating characteristic (ROC) curve as the value of quantitative MRD with maximal Youden index. Though steps (ii) and (iii) may give numerically different results, those values should be very close. Statistical analysis was done with SAS® Statistical Analysis System Software Version 9.2 (SAS Institute Inc., Cary, NC, USA) under Windows® XP-Professional on an IBM® ThinkPad® platform. For step (i), procedure PROC NPAR1WAY was used which also gives a result for step (ii). The ROC analysis of step (iii) was then done on the OUTROC data set created by PROC LOGISTIC. Results are listed in Table S1 for MRD values before and after HDT and autologous PB SCT with respect to two binary endpoints, relapse (yes/no) and death (yes/no).

Among the 53 patients who entered the study, 40 relapses and 29 deaths were observed during follow-up. Post-transplantation MRD was available from only 39 patients among which 29 relapses and 19 deaths were observed. The statistical distributions of pre- and post-transplantation MRD values were compared between relapsing and non-relapse patients as well as between follow-up survivors and non-surviving patients by nonparametric Kolmogorov-Smirnov tests. While post-transplantation MRD distributions did not differ statistically, the pre-transplantation MRD distributions did so between survivors and non-survivors at nominal significance level of P < 0.05 and – as a borderline result – between relapsing and non-relapse patients at a nominal significance level of P = 0.06 (Table S1).

Sensitivity and specificity estimate from the ROC analysis translate into proportions of patients with pre-transplantation MRD > 0.2% 2IgH/ß-actin ratio among non-survivors, 0.78 (95%ci: 0.56-0.93), and patients with pre-transplantation MRD ≤ 0.2% 2IgH/ß-actin ratio among survivors, 0.60 (95%ci: 0.40-0.78), respectively. From the Kolmogorov-Smirnov test, the threshold value of 0.12 analogously yields estimates of 0.61 (95%ci: 0.38-0.81) and 0.60 (95%ci: 0.40-0.78), respectively.

As prognostic considerations start from classification of pre-transplantation MRD, the positive and negative reverse proportions, calculated as sample prediction values, are more applicable here: the probability of later death for patients with pre-transplantation MRD > 0.2% 2IgH/ß-actin is 0.54 (95%ci: 0.33-0.74) and the probability of survival during follow-up for patients with pre-transplantation MRD ≤ 0.2% 2IgH/ß-actin is 0.67 (95%ci: 0.46-0.84). The respective results for the threshold of 0.12% 2IgH/ß-actin from the Kolmogorov-Smirnov test are 0.60 (95%ci: 0.40-0.78) and 0.78 (95%ci: 0.56-0.93) for death and survival among those surpassing and not surpassing that threshold, respectively.

In conclusion, a significant prognostic threshold level of the 2IgH/ß-actin ratio from BM samples before HDT could be defined. The use of a cut-off value of 0.2% 2IgH/ß-actin, as determined by ROC analysis, resulted in the most sensitive and specific prognostic estimate.


Table S1. Comparison of patients with and without relapse and of surviving and not surviving patients during follow-up by the respective cumulative percentage curves with a Kolmogorov-Smirnov test, MRD values of maximal distance and of maximal Youden index.

<table>
<thead>
<tr>
<th>Test for difference</th>
<th>Kolmogorov-Smirnov test</th>
<th>ROC analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P value</td>
<td>MRD</td>
</tr>
<tr>
<td></td>
<td>No. cases/total no.</td>
<td>[%IgH/β-actin]</td>
</tr>
<tr>
<td>Relapse</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MRD before HDT</td>
<td>40/53</td>
<td>0.06</td>
</tr>
<tr>
<td>MRD after HDT</td>
<td>29/39</td>
<td>0.13</td>
</tr>
<tr>
<td>Dead</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MRD before HDT</td>
<td>23/53</td>
<td>0.04</td>
</tr>
<tr>
<td>MRD after HDT</td>
<td>19/39</td>
<td>0.60</td>
</tr>
</tbody>
</table>

*Sensitivity estimate: 0.6087 (95%ci: 0.38-0.81), specificity estimate: 0.6000 (95%ci: 0.40-0.78).
†Sensitivity estimate: 0.7826 (95%ci: 0.56-0.93), specificity estimate: 0.6000 (95%ci: 0.40-0.78).
‡Values in brackets because of borderline statistical significance at 0.05 < P ≤ 0.10.
§n.a. = not applicable for more than borderline lack of statistical significance (P > 0.10).