Impact of centralized diagnostic review on quality of initial staging in Hodgkin lymphoma: experience of the German Hodgkin Study Group

Paul J. Bröckelmann,1,2 Helen Goergen,2 Michael Fuchs,1,2 Jan Kriz,2,3 Robert Seemrau,2,4 Christian Baues,2,4 Carsten Kobe,2,5 Karolin Behringer,1,2 Dennis A. Eichenauer,1,2 Bastian von Tresckow,1,2 Beate Klimm,2,6 Teresa Halbsguth,2,7 Diana Wongso,2 Annette Plutschow,2 Heinz Haverkamp,1 Markus Dietlein,2,5 Hans T. Eich,2,3 Harald Stein,2,6 Volker Diehl,2 Peter Borchmann1,2 and Andreas Engert1,2

1Department I of Internal Medicine, University Hospital of Cologne, 2German Hodgkin Study Group (GHSG), Cologne, 3Department of Radiation Oncology, University Hospital Münster, Münster, 4Department of Radiation Oncology, University Hospital of Cologne, 5Department of Nuclear Medicine, University Hospital of Cologne, Cologne, 6Department III of Internal Medicine, Krankenhaus Düren, Düren, 7Department II of Internal Medicine, University Hospital of Frankfurt, Frankfurt, and 8Pathodiagnostik Berlin, Berlin, Germany

Received 6 May 2015; accepted for publication 23 July 2015

Correspondence: Professor Dr Andreas Engert, Department I of Internal Medicine, University Hospital of Cologne, Kerpener Str. 62, D-50937 Cologne, Germany.
E-mail: a.engert@uni-koeln.de

Summary

Accurate clinical staging is crucial for adequate risk-adapted treatment in Hodgkin lymphoma (HL) to prevent patients from under- or over-treatment. Within the latest German Hodgkin Study Group trial generation, diagnostic findings such as histopathology, computerized tomography imaging and clinical risk factors were re-evaluated by expert panels. Here, we retrospectively analysed 5965 patients and identified 399 in who major discordant findings changed their first-line treatment allocation. Histopathology review did not confirm the initial diagnosis of HL in 87 patients. Treatment allocation was revised in 312 of the remaining 5878 patients: 176 were assigned to a higher and 128 to a lower risk group, respectively; the correct treatment group remained unclear in 8 patients. Cases of revised treatment allocation accounted for 9.8%, 6.0%, 0.8%, and 14.8% of patients initially assigned to the HD13, HD14, HD15 trials and stage IA lymphocyte-predominant HL project, respectively. Most revisions were due to wrong application of clinical stage (20.5% of 312 patients with revised treatment group), histological subtype (9.0%) or the risk factors ≥3 involved areas (46.8%) or large mediastinal mass (9.3%). In conclusion, centralized review by experienced experts changed risk-adapted first-line treatment in a relevant proportion of HL patients. Quality control measures clearly improve the accuracy of treatment and should be implemented in clinical practice.

Keywords: Hodgkin lymphoma, staging, quality control, accuracy, risk-adapted therapy.

Despite substantial progress in the treatment of Hodgkin lymphoma (HL) (Engert et al, 2012; von Tresckow et al, 2012; Behringer et al, 2014), therapy-related side effects, such as secondary malignancies, organ toxicities, infertility, fatigue and impaired quality of life are of major concern (Aleman et al, 2003; Heute et al, 2009; Swerdlow et al, 2011; Sasse et al, 2012; Behringer et al, 2013). Recent activities in this field thus focus on more individualized therapeutic concepts to further reduce treatment-related side effects while maintaining the current outcome levels (Hay et al, 2013; Klimm et al, 2013). To implement truly customized therapeutic approaches that prevent patients from being under- or over-treated, correct treatment group allocation (TGA) based on detailed assessment of histopathological profile and potential targetable structures is essential (Eichenauer et al, 2014).

Discrepancies in the interpretation of diagnostic findings between local institutions and centralized review with impact on TGA and radiotherapy (RT) volume have been observed in a relevant proportion of patients within several studies (Eich et al, 2004; Stevens et al, 2012). Within the German Hodgkin Study Group (GHSG) HD13, HD14, HD15 trials...
and the observational project for stage IA lymphocyte-predominant HL (NLPHL-IA), diagnostic findings, including histopathology and images from contrast-enhanced computerized tomography (CT) scans, were centrally reviewed by GHSG multidisciplinary quality panels. To better understand the impact of our quality control measures for initial staging we analysed the proportion of patients with revised histopathological diagnosis or treatment group re-allocation according to the GHSG risk stratification.

Methods

Patients

Between 28 January 2003 and 30 September 2009, over 400 trial centres in Germany, Czech Republic, Switzerland, Austria, and the Netherlands recruited patients for the concurrent GSHG trials HD13 ($n = 1694$), HD14 ($n = 2032$), HD15 ($n = 2173$) and NLPHL-IA ($n = 81$). Patients with newly diagnosed HL were allocated to one of the trials according to their treatment group as described below. All patients provided written informed consent before study entry. Fifteen patients withdrew consent and were thus excluded, resulting in a total of 5965 analysed patients.

Patients received risk-adapted treatment consisting of multi-agent chemotherapy, RT or both modalities. Treatment groups were defined by clinical stage and the presence of risk factors including (a) large mediastinal mass (≥1/3 of the maximal thoracic diameter), (b) extranodal disease, (c) elevated erythrocyte sedimentation rate (≥50 mm/h without B symptoms, ≥30 mm/h with B symptoms), and (d) involvement of three or more nodal areas. Patients with early-stage favourable HL, i.e. those in clinical stage I–II without risk factors (a–d), qualified for the HD13 trial, except for patients diagnosed with stage IA nodular lymphocyte-predominant HL (NLPHL-IA). Patients with stage I–II and at least one of the risk factors (a–d) were classified as early-stage unfavourable HL and qualified for the HD14 trial, except for patients having stage IIB and at least one of the risk factors (a–b). These patients, as well as those in clinical stage III–IV, qualified for the HD15 trial for advanced-stage HL. Except for the NLPHL-IA, all trials were registered at Current Controlled Trials (ISRCTN63474366, ISRCTN04761296, ISRCTN32443041) and recently published (Engert et al., 2012; von Tresckow et al., 2012; Behringer et al., 2014). In NLPHL-IA, patients were treated with 30 Gy involved-field RT (IFRT) and followed up for tumour status and survival (Eichenauer 2015).

The procedures for diagnosis, staging, and trial allocation were similar in all studies. There was no screening phase and patients were randomized without prior centralized review. Patients for whom HL was diagnosed by a primary care pathologist had their tissue samples sent to one of six reference pathology review centres within the GHSG network. The patient was removed from the trial analysis set if HL diagnosis was disconfirmed.

Initial clinical staging was conducted by the treating physician and included contrast-enhanced CT imaging performed and interpreted locally. All diagnostic findings were reported to the GHSG Trial Coordination Centre (TCC). Trial allocation followed by randomization was then performed within the GHSG database. After randomization, a GHSG trial physician reviewed disease extent documentation. In addition, for all patients enrolled into HD13, HD14 and NLPHL-IA, initial CT scans were sent to the University Hospital of Cologne, where a panel of experts in imaging and radiation oncology re-evaluated the initial findings and developed an IFRT plan for each individual patient. In the HD15 trial, initial imaging was only reviewed in patients with a residual mass ≥2.5 cm after completion of chemotherapy ($\approx 1/3$ of patients), because the protocol treatment did not routinely include IFRT (Engert et al., 2012).

Incorrect allocation to a trial either revealed by the GHSG trial physician or the imaging panel, or based on a correction of initial findings reported to the TCC by the local physician, was documented in the trial database. If protocol treatment had not been started by then, it was at that time still possible to remove the patient from the original trial and register them for the appropriate trial. If the correction was found after treatment had been initiated, the treating physician was informed immediately to stop protocol treatment and treat the patient according to the correct risk group.

In preparation for interim and final analyses, comprehensive data checks were performed by the TCC. Some additional cases with incorrect TGA were identified during this procedure, which were mainly stage IA patients assigned to HD13 that were diagnosed with NLPHL by histopathological review. However, given that checks were mostly performed for patients that already finished protocol treatment, therapy could not be adjusted.

Statistical analysis

Patients from HD13, HD14, HD15 and NLPHL-IA with documented correction of the initial histopathological diagnosis or TGA were identified within the GHSG database and analysed descriptively for details regarding the correction, including initial and corrected treatment group, decisive corrections and localizations, and time to correction. Patients enrolled into a second GHSG trial after revision of TGA as described above were here analysed according to the trial they were incorrectly first registered to.

Potential risk factors for up or down revision of TGA were assessed by performing multivariate logistic regression analyses; $P$-values <0.05 were considered significant. Patients initially assigned to NLPHL-IA were excluded from the model evaluating down-staging, and patients assigned to HD15 were excluded from the model for up-staging.

Statistical analyses were performed using SAS software (version 9.4 for Microsoft Windows; SAS Institute, Cary, NC, USA).
Results

Of the 5965 analysed cases, 87 were removed from the trials' analysis sets based on disconfirmation of HL diagnosis. Of the remaining 5878 patients (see Table I for patient characteristics), 312 had been initially assigned to an inappropriate treatment group (Fig 1). Of these patients, 176 (56.4%) were then upstaged and 128 (41.0%) were downstaged while the correct risk group remained unclear due to insufficient diagnostics or documentation in eight patients (2.6%). The proportion of HL patients initially assigned to an inappropriate treatment group was 9.8% in HD13, 6.0% in HD14, 0.8% in HD15 and 14.8% in NLPHL-IA.

Histopathology

Histopathology review was documented for 5269 patients (88.3%) and accounted for disconfirmation of HL diagnosis in 87 of them (1.7%). In most cases a non-HL was diagnosed (70 cases), but expert review also revealed no malignant disease in 17 patients, of whom the majority had initially been allocated to the early-stage favourable HL risk group.

Additionally, correction of histological subtype from classical HL to NLPHL or vice versa resulted in different TGA

Table I. Baseline patient demographics and clinical characteristics of HL patients according to study.

|                  | NLPHL-IA (n = 81) | HD13 (n = 1662) | HD14 (n = 2009) | HD15 (n = 2126) | Total* (n = 5878) |
|------------------|-------------------|-----------------|-----------------|-----------------|-------------------|
|                  | N                 | %               | N               | %               | N                 | %               |
| Age, years: median (range) | 38 (19–75) | 39 (18–75) | 32 (18–61) | 33 (18–60) | 35 (18–75) |
| Male              | 62                | 76-5            | 982             | 59-1            | 1053             | 52-4            | 836             | 39-3            | 3290             | 56-0            |
| Female            | 19                | 23-5            | 680             | 40-9            | 1053             | 52-4            | 836             | 39-3            | 2588             | 44-0            |
| Country           |                   |                 |                 |                 |                   |                 |                 |                 |                   |                 |
| Germany           | 78                | 96-3            | 1478            | 88-9            | 1662             | 82-7            | 1808            | 85-0            | 5026             | 85-5            |
| Czech Republic    | 3                 | 3-7             | 44              | 2-6             | 180              | 9-0             | 163             | 7-7             | 390              | 6-6             |
| Switzerland       | 7                 | 8-8             | 88              | 4-4             | 67               | 3-2             | 226             | 3-8             |                   |                 |
| Austria           | 3                 | 8-4             | 49              | 3-2             | 34               | 1-6             | 121             | 1-0             | 115              | 2-0             |
| The Netherlands   | 3                 | 1-9             | 30              | 1-5             | 54               | 2-5             |                   |                 |                   |                 |
| Centre type       |                   |                 |                 |                 |                   |                 |                 |                 |                   |                 |
| University hospital | 32              | 39-5            | 537             | 32-3            | 808              | 40-2            | 935             | 44-0            | 2312             | 39-3            |
| Other hospital    | 27                | 33-3            | 679             | 40-9            | 752              | 37-4            | 833             | 39-2            | 2291             | 39-0            |
| Practice          | 14                | 17-3            | 270             | 16-2            | 274              | 13-6            | 209             | 9-8             | 767              | 13-0            |
| Specialised practice | 6             | 7-4             | 171             | 10-3            | 173              | 8-6             | 146             | 6-9             | 496              | 8-4             |
| Radiotherapeutic clinic or practice | 2 | 2-5 | 5 | 0-3 | 3 | 0 | 3 | 0 | 12 | 0-2 |
| Clinical stage (according to data recorded at registration)† | | | | | | | | | | |
| IA               | 81                | 100-0           | 670/1659        | 40-4            | 124              | 6-2             | 364             | 17-1            | 875/5875         | 14-9            |
| IB               |                   |                 | 54/1659         | 3-3             | 56               | 2-8             | 110/5875        | 1-9             |                   |                 |
| IIA              |                   |                 | 855/1659        | 51-5            | 1307             | 65-1            | 2162/5875       | 36-8            |                   |                 |
| IIB              |                   |                 | 80/1659         | 4-8             | 522              | 26-0            | 875/5875        | 16-4            |                   |                 |
| IIIA             |                   |                 | 585             | 27-5            | 585/5875         | 10-0            |                   |                 |                   |                 |
| IIIB             |                   |                 | 564             | 26-5            | 564/5875         | 9-6             |                   |                 |                   |                 |
| IVA              |                   |                 | 198             | 9-3             | 198/5875         | 3-4             |                   |                 |                   |                 |
| IVB              |                   |                 | 415             | 19-5            | 415/5875         | 7-1             |                   |                 |                   |                 |
| Risk factors (according to data recorded at registration)‡ | | | | | | | | | | |
| Large mediastinal mass‡ | 344 | 17-1 | 595 | 28-0 | 939 | 16-0 | | | | |
| Extranodal disease | 124 | 6-2 | 350 | 16-5 | 474 | 8-1 | | | | |
| Three or more nodal areas | 1326 | 66-0 | 1727 | 81-2 | 3053 | 51-9 | | | | |
| Elevated ESR     | 1007             | 50-1            | 1279            | 60-2            | 2286             | 38-9            |                   |                 |                   |                 |
| Correction of initial treatment group | | | | | | | | | | |
| No               | 69                | 83-2            | 1499            | 90-2            | 1889             | 94-0            | 2109            | 99-2            | 5566             | 94-7            |
| Yes              | 12                | 14-8            | 163             | 9-8             | 120              | 6-0             | 17              | 0-8             | 312              | 5-3             |

HL, Hodgkin lymphoma; NLPHL-IA, stage IA nodular lymphocyte-predominant Hodgkin lymphoma; ESR, erythrocyte sedimentation rate.

*87 cases with disconfirmation of histopathological HL diagnosis excluded.

†Information on B-symptoms missing in 3 patients.

‡At least one-third of maximal thorax diameter.
for 26 HD13 and two NLPHL-IA patients in stage IA without any risk factors and thus accounted for 9-0% of corrections of TGA (Table II). In HD13, more males than females were down-staged due to histological subtype [21 of 982 male (2.1%) and 5 of 680 female (0.7%) patients].

**Extent of disease and risk factors**

The review of diagnostic findings, including contrast-enhanced CT scans for initial staging, resulted in a correction of TGA in 254 patients and thus accounted for 81.4% of corrections of TGA. In particular, review of imaging led to the correction of clinical stage in 64 patients (20.5% of corrections of TGA) and correction of the GHSG risk factors ≥ 3 nodal areas in 146 (46.8%), large mediastinal mass in 29 (9.3%), and extranodal involvement in 15 patients (4.8%), respectively (Table II).

Of those patients with different TGA due to correction of clinical stage, 26 switched from stage I or II to stage III or IV and 11 vice versa, representing a major change in disease extent. Critical nodal localizations were mostly found below the diaphragm and more often resulted in down-staging, whereas upstaging of patients was frequently caused by correction of organ involvement (Table SII). In the remaining 27 patients, changes from stage I to stage II or vice versa accounted for altered TGA due to the simultaneous presence of risk factors or NLPHL histology.

Decisive locations for down-staging due to rejection of the risk factor ≥ 3 nodal areas were usually localized above the diaphragm in both male and female patients; they were found across all supradiaphragmatic regions (Fig 2).

Upstaging due to this risk factor was mostly related to supradiaphragmatic lymph nodes in women; in men, upstaging was often due to additional lesions detected below the diaphragm (Fig 2).

Correction of the risk factor extranodal involvement, nearly exclusively resulted in upstaging and was caused by infiltration of varying localizations, such as thoracic wall, nasopharyngeal tissue, bones or lung.

**Risk factors for incorrect TGA**

Significant risk factors for downstaging included male sex [odds ratio (OR) 1.61; P = 0.02] and the trial the patient was initially assigned to (OR 2.91 for HD14 versus HD13 and 0.59 for HD15 versus HD13; P < 0.0001). A non-significant trend towards an increased risk for downstaging was observed in patients of older age and higher body mass index (Table III).

Risk of upstaging was significantly decreased in male patients (OR 0.71; P = 0.03) and older patients (OR 0.98; P < 0.0001), and increased in patients with infradiaphragmatic disease (OR 1.99; P = 0.005). The initial trial also had a significant impact (OR 0.23 for HD14 versus HD13 and 1.94 for NLPHL-IA versus HD13; P < 0.0001) (Table III).

**Time until correction & treatment**

Median time from enrolment to correction of TGA was 1-6 months (range 0-0–121-1; Table II). Detection of incorrect TGA took longer when the histological subtype was correction (median 38-1 months, range 0-0–110-0). Correction
Table II. Details on corrections of treatment group according to study.

| Corrected treatment group | NLPHL-IA (n = 12) | HD13 (n = 163) | HD14 (n = 120) | HD15 (n = 17) | Total (n = 312) |
|---------------------------|-------------------|----------------|----------------|--------------|----------------|
|                           | N     | %    | N     | %    | N     | %    | N     | %    | N     | %    |
| NLPHL-IA                  | 36    | 22.1 | 74    | 61.7 | 14    | 82.4 | 52    | 16.7 | 37    | 11.9 |
| Early-stage favourable HL | 10    | 83.3 |        |      | 3     | 17.6 | 87    | 27.9 |        |      |
| Early-stage unfavourable HL | 1    | 83.3 | 113   | 69.3 | 14    | 82.4 | 128   | 41.0 |        |      |
| Advanced-stage HL         | 10    | 6.7  | 42    | 35.0 | 25    | 15.2 | 52    | 16.7 |        |      |
| Unclear                   | 1     | 8.3  | 4     | 2.5  | 3     | 2.5  | 8     | 2.6  |        |      |
| Type of correction         |        |      |        |      |        |      |        |      |        |      |
| Correction towards lower stage | 36    | 22.1 | 75    | 62.5 | 17    | 100  | 128   | 41.0 |        |      |
| Correction towards higher stage | 11   | 91.7 | 123   | 75.5 | 42    | 35.0 | 176   | 56.4 |        |      |
| Unclear                   | 1     | 8.3  | 4     | 2.5  | 3     | 2.5  | 8     | 2.6  |        |      |
| Months from registration to | 2.7   |     (0.3–22.9) | 1.5   | (0.0–121.1) | 1.8   | (0.0–116.2) | 0.5   | (0.0–86.5) | 1.6   | (0.0–121.1) |
| cancellation from study, median (range) |        |      |        |      |        |      |        |      |        |      |
| Decisive correction        |        |      |        |      |        |      |        |      |        |      |
| Correction due to          |        |      |        |      |        |      |        |      |        |      |
| Clinical stage             | 8     | 66.7 | 17    | 10.4 | 28    | 23.3 | 11    | 64.7 | 64    | 20.5 |
| RF Large mediastinal mass  | 10    | 6.1  | 16    | 13.3 | 3     | 17.6 | 29    | 9.3  |        |      |
| RF Extranodal disease      | 1     | 8.3  | 4     | 3.3  | 1     | 5.9  | 15    | 4.8  |        |      |
| RF Three or more nodal areas | 89   | 54.6 | 57    | 47.5 | 146   | 46.8 |        |      |        |      |
| RF Elevated ESR           | 7     | 4.3  | 8     | 6.7  | 15    | 4.8  |        |      |        |      |
| B-symptoms                | 1     | 0.6  | 4     | 3.3  | 2     | 11.8 | 7     | 2.2  |        |      |
| Histologic subtype        | 2     | 16.7 | 26    | 16.0 | 28    | 9.0  |        |      |        |      |
| Insufficient diagnostics  | 4     | 2.5  | 3     | 2.5  | 7     | 2.2  |        |      |        |      |
| Unclear                   | 1     | 8.3  | 4     | 2.5  | 3     | 2.5  | 8     | 2.6  |        |      |
| Treatment received*       | 5/9   | 55.6 | 43/158 | 27.2 | 19/117 | 16.2 | 8    | 47.1 | 75/501 | 24.9 |
| Whole treatment according to | 120   | 87%  |        | 117   | 65%  |        | 17   | 56%  |        | 312  |
| correct stage             | 19/117 | 16.2 | 8    | 47.1 | 75/501 | 24.9 |        |      |        |      |
| Adjustment towards correct stage | 63/158 | 39.9 | 39/117 | 33.3 | 4  | 23.5 | 106/501 | 35.2 |
| Undertreatment             | 4/9   | 44.4 | 28/158 | 17.7 | 14/117 | 12.0 | 110   | 35.3 |        |      |
| Overtreatment              | 24/158 | 15.2 | 45/117 | 38.5 | 5   | 29.4 | 74/501 | 24.6 |        |      |

HL, Hodgkin lymphoma; NLPHL-IA, nodular lymphocyte-predominant Hodgkin lymphoma; RF, risk factor; ESR, erythrocyte sedimentation rate. *Information on treatment missing in 11 patients.

The accuracy of RT has been studied in detail (Bijker et al, 2001; Kouloulias et al, 2003; Eich et al, 2004; Fairchild et al, 2012; Kriz et al, 2012), less is known regarding the overall quality of initial staging of HL patients. We thus analysed the effect of our centralized quality control measures on correct allocation to a treatment group in a total of 5965 HL patients treated in the GHSG trials HD13, HD14, HD15 and NLPHL-IA. We found that centralized, multidisciplinary review of diagnostic procedures clearly changed the recommended treatment in a relevant proportion of patients. Revision of the initial treatment group was most frequently caused by a different histopathological diagnosis or revised interpretation of CT imaging.

In the present analysis, 87 patients (1.7%) had a different histology upon expert pathology revision, including no malignant disease in 17 patients and a different lymphoma in 70 patients. Overall, these numbers are substantially lower than observed in an earlier analysis reflecting implementation of consistent classifications and immunohistochemical staining in malignant lymphoma: Whereas 8.3% of cases were revised in 1993 (Georgii et al, 1993), this figure was down to...
Fig 2. Decisive localizations in patients with correction of the risk factor '≥3 nodal areas'. Downward or upward correction denotes an adjustment towards a more favourable (A, B) or unfavourable (C, D) treatment group, respectively. (A) Male patients with downward correction*. (B) Female patients with downward correction†. (C) Male patients with upward correction‡. (D) Female patients with upward correction§. *Including 14 patients with corrections in two areas and two patients with corrections in three areas. †Including three patients with corrections in two areas and one patient with corrections in four areas. ‡Including eight patients with corrections in two areas, one patient with corrections in four areas and one patient where corrected localizations are unknown. §Including 10 patients with corrections in two areas, one patient with corrections in three areas and one patient where corrected localizations are unknown.
Table III. Multivariate analysis for correction of treatment group.

| Study | Odds ratio [95% confidence interval] | p |
|-------|-------------------------------------|---|
| HD13 [reference group] | 2.91 [1.80–4.70] | <0.0001 |
| HD14 | 0.59 [0.26–1.34] | 0.21 |
| Sex [male versus female] | 1.61 [1.08–2.41] | 0.02 |
| Age at enrolment [years] | 1.02 [1.00–1.03] | 0.08 |
| Body mass index | 1.05 [0.99–1.07] | 0.13 |
| Centre type | 1.00 |
| University hospital [reference group] | 1.00 |
| Other hospital | 1.00 |
| Specialized practice | 1.05 [0.52–2.12] | 0.89 |
| General practice | 1.06 [0.59–1.90] | 0.85 |
| Infradiaphragmatic involvement [yes versus no] | 0.75 [0.38–1.47] | 0.40 |

Multivariate analysis for upward correction of treatment group†

| Study | Odds ratio [95% confidence interval] | p |
|-------|-------------------------------------|---|
| HD13 [reference group] | 0.23 [0.16–0.33] | <0.0001 |
| NLPHL-IA | 1.94 [0.98–3.84] | 0.06 |
| Sex [male versus female] | 0.71 [0.52–0.97] | 0.03 |
| Age at enrolment [years] | 0.98 [0.96–0.99] | <0.0001 |
| Centre type | 0.73 |
| University hospital [reference group] | 0.99 [0.69–1.43] | 0.96 |
| Other hospital | 0.99 [0.72–1.43] | 0.96 |
| Specialized practice | 1.12 [0.64–1.94] | 0.69 |
| General practice | 1.25 [0.80–1.96] | 0.33 |
| Infradiaphragmatic involvement [yes versus no] | 1.99 [1.23–3.22] | 0.005 |
| Body mass index‡ | 0.98 [0.94–1.02] | 0.26 |

Downward or upward correction denotes an adjustment towards a more favourable or unfavourable treatment group, respectively. NLPHL-IA, nodular lymphocyte-predominant Hodgkin lymphoma. *Multivariate logistic regression model including 5564 Hodgkin lymphoma patients (cases with disconfirmation of histopathological diagnosis, NLPHL-IA and patients with missing values excluded). †Multivariate logistic regression model including 3735 Hodgkin lymphoma patients (cases with disconfirmation of histopathological diagnosis, HD15 trial and patients with missing values excluded). ‡Since documentation of height and weight was missing in a large number of patients including all patients from NLPHL-IA, body mass index was not included in the primary model. Displayed results for body mass index emerge from a separate model including 3498 patients with available documentation of height and weight and all factors from the primary analysis in addition to body mass index.

3.0% in 2001 (Glaser et al, 2001). More recently, a report on expert review within a UK network on various lymphoid malignancies revealed discordant histopathological findings relevant for TGA in 2.1% of cases (Proctor et al, 2011). The distinction between different disease entities is highly relevant to every single patient given that treatment approaches may differ fundamentally, especially in those patients not suffering from any malignant disease. Reliable identification of targetable structures expressed within the tumour additionally represents a challenging task in the era of targeted therapies.

Corrections of clinical stage or the risk factor ≥3 nodal areas accounted for the majority of cases re-allocated due to revised CT imaging. Of note, downstaging due to correction of clinical stage might be underreported in our analysis, because initial imaging was not reviewed systematically in HD15. Relevant changes in clinical stage were mainly due to additionally detected organ involvement or absence of infradiaphragmatic lesions. In contrast, a change of the risk factor ≥3 nodal areas mainly occurred in the cervico-thoracic junction and paraaortic region. Discrimination of adjacent lymph nodes and assignment to one of the predefined areas seems to be difficult in these regions of complex and dense anatomic structures and hence constitutes a well known challenge to clinicians (De-chow et al, 2002). Uncertainty regarding the exact definition of this by now well-established GHSG risk factor (Sieber et al, 2002) and distinction of the different areas may account for some of the revisions.

To define a subgroup of patients at risk for incorrect TGA, we performed multivariate analyses for potential risk factors. These analyses showed that initial treatment group, age, sex and infradiaphragmatic lesions played a decisive role. Interestingly, centre type did not affect the risk of incorrect TGA. This is in accordance with a previous analysis by the GHSG that found no difference in the treatment outcome with regard to centre type or size (Klimm et al, 2012). The GHSG quality control programme as well as prompt consulting by study physicians via telephone in case of uncertainties prior to enrolment may contribute to this observation.

The quality of initial staging has been addressed in other analyses with a focus on RT, as successful reduction of the RT fields requires precise staging (Girinsky et al, 2006). As previously reported (Kria et al, 2012), centralized review revealed minor corrections of disease involvement in 56% and 72% of patients allocated to HD13 and HD14, respectively, which are also included in the present analysis. These corrections resulted in revised IFRT fields in 38%, which also contributes to the adequate application of risk-adapted therapy. A more recent analysis in 124 adult HL patients treated at Dutch hospitals outside clinical trials reported necessary corrections of treatment including RT in 19% of cases (Stevens et al, 2012). A similar analysis in 578 paediatric HL patients revealed incorrect staging in 20% of patients, resulting in a correction of treatment group in 13% of patients (Dieckmann et al, 2002). Centralized review by a paediatric Hodgkin study group had to be omitted for 2 years due to insufficient funding which allowed Luders et al (2014) to compare 142 paediatric patients treated outside clinical trials to those treated within the GPOH-HD95 trial. Incorrect TGA and RT-fields resulted in inadequate treatment in 39% of patients leading to a significantly poorer progression-free survival (PFS) in this group of patients (Luders et al, 2014).
In most patients included in our analysis, timely adjustment of treatment was possible due to prompt review of diagnostic findings. Given that all review procedures, including shipment of imaging and histopathological material, were only carried out after registration and start of treatment in our studies, time until correction might be further reduced by implementation of a screening phase prior to enrolment. Interestingly, we did not observe a sequential improvement in the quality of TGA over the course of the study durations. This might reflect the existing challenge posed by certain aspects of initial staging of HL and emphasizes the need for additional modalities and/or expert review. All diagnostic findings should be interpreted carefully, especially in patients with limited stage or infradiaphragmatic disease, to facilitate adequate TGA also without centralized review.

As a potential limitation of our analysis, we were not able to assess the impact of incorrect TGA on PFS or overall survival (OS). The group of patients at risk for impaired outcome comprises those potentially undertreated due to an upstaging without adjustment of treatment. Due to a timely correction of risk group in the majority of cases, this applied to only 46 patients in our analysis. As some of those patients responded well to the initial treatment approach, additional treatment according to the revised risk group might not have been deemed necessary by the treating physician. Furthermore, due to incomplete documentation after revision of risk group and exclusion from the trial, it remains unclear whether some of these patients were actually undertreated. Thus, it is not possible to draw sound conclusions regarding PFS or OS from this small and heterogeneous subset.

From an oncologist’s perspective, correct TGA is especially important when pursuing modern risk-adapted approaches to ensure high quality of care. It is also essential for patients treated within clinical trials to help improving treatment strategies based on data collected in such a trial. Differentiation of risk groups has demonstrated significant impact on PFS and OS, even in early-stage HL (Klimm et al., 2013). Hence, inappropriate treatment may result in refractory disease and premature death, which is a disaster for the individual patient and results in a loss of productivity in the society (Hanly et al., 2015). Whereas costs for centralized review of staging images that were already generated is difficult to measure, the costs for detection of one patient with histopathological misdiagnosis of malignant lymphoma in the UK were 690€ (Proctor et al., 2011). In comparison to the relevant socioeconomic burden in case of relapsed disease or premature death, reasonable costs for centralized review in all HL patients could hence be a useful investment, also for treatment outside clinical trials and with no sponsored funding.

Furthermore, the routine use of additional modalities with high sensitivity and specificity, such as a $^{18}$F-fluorodeoxyglucose-positron emission tomography ($^{18}$F-FDG-PET) scan (Isasi et al., 2005; Cheson et al., 2014), which frequently leads to a correction of disease involvement (Moog et al., 1998; Jerusalém et al., 2001), could improve the quality of initial staging. Despite increased sensitivity, correct interpretation and assessment of PET scans is challenging due to false-positive lesions and heterogeneity in criteria: Whereas divergence is high among observers with varying experience (Zijlstra et al., 2007), inter-observer agreement seems to be high in experienced physicians (Hofman et al., 2009) using pre-defined guidelines (Barrington et al., 2010). These findings reflect the importance of qualified evaluation and consistent classifications in any type of imaging.

In summary, we identified 399 cases (6.7%) of incorrect diagnosis or staging resulting in major changes to first-line treatment among 5965 patients registered for clinical HL trials in this largest analysis on quality assessment of initial staging conducted so far. Given that exact initial staging of HL is crucial for risk-adapted treatment, multidisciplinary expert review and more sensitive modalities, such as PET, should be implemented in initial staging for all HL patients to ensure high quality of care. Within clinical trials, a screening phase and centralized review could be implemented to assure high-quality data.

Acknowledgements

The HD13, HD14 and HD15 trials have been funded by Deutsche Krebshilfe and the Swiss Federal Government. We are grateful to all patients and colleagues who participated in our HD13, HD14 and HD15 trials and the NLPHL-IA observational project.

Author contributions

PJB, MF, KB, DE, BvT, BK, TH and DW reviewed diagnostic findings at the GHSG TCC. JK, RS, CB, CK, MD and HTE undertook the central review of the initial imaging. HS coordinated the reference pathology. MF coordinated activities at the GHSG TCC. HG, AP and HH were the lead statisticians for the HD13, HD14 and HD15 trials, respectively, which were designed by VD, PB and AE. PJB, HG and AE designed this analysis and wrote the first draft of the article. HG did the statistical analyses and all authors contributed to data interpretation, reviewed the draft and approved the final version.

Competing interests

The authors have no competing interests.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1. Decisive correction according to study and type of correction.

Table SII. Decisive localizations in patients with correction of clinical stage from I/II to III/IV or vice versa.
Impact of risk factors on outcomes in early-stage Hodgkin’s lymphoma: an analysis of international staging definitions. *Annals of Oncology*, 24, 3070–3076.

Kouloulias, V.E., Poortmans, P.M., Bernier, J., Heriot, J.C., Johansson, K.A., Davis, B., Godson, F., Garavaglia, G., Pierart, M. & van der Schueren, E. (2003) The Quality Assurance programme of the Radiotherapy Group of the European Organization for Research and Treatment of Cancer (EORTC): a critical appraisal of 20 years of continuous efforts. *European Journal of Cancer*, 39, 430–437.

Kriz, J., Bangard, C., Haverkamp, U., Bongartz, R., Baues, C., Engert, A., Mueller, R.P. & Eich, H.T. (2012) Quality control of involved-field radiotherapy for patients with early stage Hodgkin’s lymphoma based on a central prospective review. Comparison of the results between two study generations of the German Hodgkin Study Group. *Strahlentherapie und Onkologie*, 188, 660–665.

Luders, H., Ruhl, U., Marciniak, H., Haerting, J., Clavier, A., Mann, G., Schellong, G., Wickmann, L. & Dorffel, W. (2014) The impact of central review and central therapy planning on the treatment of children and adolescents with Hodgkin lymphoma. *European Journal of Cancer*, 50, 425–433.

Moog, F., Bangarter, M., Diederichs, C.G., Gahrmann, A., Merkle, E., Frickhofen, N. & Reske, S.N. (1998) Extranodal malignant lymphoma: detection with FDG PET versus CT. *Radiology*, 206, 475–481.

Proctor, I.E., McNamara, C., Rodriguez-Justo, M., Isaacson, P.G. & Ramsay, A. (2011) Importance of expert central review in the diagnosis of lymphoid malignancies in a regional cancer network. *Journal of Clinical Oncology*, 29, 1431–1435.

Sasse, S., Klimm, B., Goergen, H., Fuchs, M., Heyden-Homerkamp, A., Lohri, A., Koch, O., Wilhelm, M., Trenn, G., Finke, J., Muller, R.P., Diehl, V., Eich, H.T., Borchmann, P. & Engert, A. (2012) Comparing long-term toxicity and efficacy of combined modality treatment including extended- or involved-field radiotherapy in early-stage Hodgkin’s lymphoma. *Annals of Oncology*, 23, 2953–2959.

Sieber, M., Tesch, H., Pfistner, B., Rueffer, U., Lathan, B., Brosteau, O., Paulus, U., Koch, T., Pfreundschuh, M., Loefller, M., Engert, A., Joesting, A., Wolf, J., Hasenclever, D., Franklin, J., Duestnek, E., Georgi, A., Schalk, K.P., Kirchner, H., Doelken, G., Munker, R., Koch, P., Hermann, R., Greil, R., Anselmo, A.P. & Diehl, V. (2002) Rapidly alternating COPP/ABV/IMEP is not superior to conventional COPP/ABVD in combination with extended-field radiotherapy in intermediate-stage Hodgkin’s lymphoma: final results of the German Hodgkin’s Lymphoma Study Group Trial HD5. *Journal of Clinical Oncology*, 20, 476–484.

Stevens, W.B., van Krieken, J.H., Mus, R.D., Arens, A.L., Mattijsen, V., Oosterveld, M., de Kruijf, E.J., de Vries, F., Koster, A., van der Maazen, R. & Raemaekers, J. (2012) Centralised multidisciplinary re-evaluation of diagnostic procedures in patients with newly diagnosed Hodgkin lymphoma. *Annals of Oncology*, 23, 2676–2681.

Swerdlov, A.J., Higgins, C.D., Smith, P., Cunningham, D., Hancock, B.W., Horwich, A., Hoskin, P.J., Lister, T.A., Radford, J.A., Rohatiner, A.Z. & Linch, D.C. (2011) Second cancer risk after chemotherapy for Hodgkin’s lymphoma: a collaborative British cohort study. *Journal of Clinical Oncology*, 29, 4096–4104.

von Tresckow, B., Plutschow, A., Fuchs, M., Klimm, B., Markova, J., Lohri, A., Kral, Z., Greil, R., Topp, M.S., Meisner, J., Zijlstra, J.M., Soeldner, M., Stein, H., Eich, H.T., Mueller, R.P., Diehl, V., Borchmann, P. & Engert, A. (2012) Dose-intensification in early unfavorable Hodgkin’s lymphoma: final analysis of the German Hodgkin Study Group HD14 trial. *Journal of Clinical Oncology*, 30, 907–913.

Zijlstra, J.M., Comans, E.F., van Lingen, A., Hoekstra, O.S., Gundy, C.M., Willem Coebergh, J. & Bongers, V. (2007) FDG PET in lymphoma: the need for standardization of interpretation. An observer variation study. *Nuclear Medicine Communications*, 28, 798–803.