Patients’ and physicians’ roles in detecting recurrent Hodgkin lymphoma following complete remission†

A. Bestawros1*, L. Foltz2, N. Srour3, K. J. Savage1 & J. M. Connors2

Departments of 1Medical Oncology; 2Hematology, British Columbia Cancer Agency Centre for Lymphoid Cancer, Vancouver; 3Department of Medicine, Ottawa Hospital Research Institute, Ottawa, Canada

Received 23 July 2012; revised 13 October 2012; accepted 16 October 2012

Background: Optimal post-treatment surveillance for patients with Hodgkin lymphoma in first complete remission (CR) is unknown. Guidelines are based on consensus rather than high-quality evidence. It is unknown if routine screening leads to earlier relapse detection or translates into better outcomes.

Patients and methods: We identified 258 patients with relapse after CR and determined whether the recurrence was detected as a result of patient-detected symptoms (PT group) or through exams or tests ordered by the physician in the absence of symptoms (MD group).

Results: Of 258 recurrences, 182 (71%) were in the PT group. The median time to diagnosis of recurrence was similar in both groups (PT group = 1.65 years; MD group = 1.95 years; P = 0.69). Neither the postrelapse progression-free (PFS, P = 0.26) nor overall survival (OS, P = 0.40) differed significantly between the groups.

Conclusion: Patients are much more likely to detect recurrence than their physicians employing routine follow-up testing. There is no difference in PFS or OS between patients whose recurrence was self-detected versus those whose recurrence is diagnosed by physician through routine screening. We found no benefit for detection of HL recurrence in asymptomatic patients and thus cannot support the routine use of costly, anxiety-provoking or potentially harmful tests in the absence of symptoms.

Key words: follow-up, Hodgkin lymphoma, recurrence

Introduction

Today the prognosis of Hodgkin lymphoma (HL) constitutes one of the best in oncology. As of 2007, the 5-year overall survival is 86% [1] despite a recurrence risk of 10%–15% for early-stage disease and up to 30% for advanced disease [2]. These favorable overall outcomes are explained not only by effective first-line treatments, but also availability of successful secondary interventions for recurrent disease employing high-dose therapy and autologous stem-cell transplantation (ASCT). The latter has an overall cure rate of 50%–60% [2]. Advanced disease stage (stage III and IV) at relapse has been identified as a negative prognostic factor for ASCT [3, 4]. However, several questions remain: Does routine testing in the absence of symptoms lead to earlier diagnosis of recurrent disease? Does

---

1*Correspondence to: Dr A. Bestawros, Department of Medicine, Division of Hematology and Oncology, Centre Hospitalier de l’Université de Montréal, Hôpital Notre-Dame - Pavillon Deschamps (8-6149), 1560 Sherbrooke Street East, Montreal, QC H2L 4M1, Canada. Tel: +1-514-869-8000 (27012); Fax: +1-514-412-7572; E-mail: alain.bestawros@umontreal.ca

1†A preliminary analysis of the results reported in this paper was presented at the meeting of the American Society of Clinical Oncology in Chicago, Illinois in June 2011.
this improve outcomes? And if so, what is the optimal follow-up of patients who have achieved a first complete remission (CR)?

These questions have yet to be answered in a prospective and definitive way. Existing guidelines are based on consensus and speculation [5–8]. With regards to the role of imaging in the follow-up after completion of therapy, the National Comprehensive Cancer Network (NCCN) recommends chest imaging [chest X-ray or computed tomography (CT)] every 6–12 months during the first 2–3 years and an abdominal/pelvic CT every 6–12 months for the first 2–3 years [5].

A few retrospective trials have examined this question. Radford et al. examined the effectiveness of routine clinical review in detecting relapsed disease and reported that 81% of relapses were discovered by the patient [9]. Routine use of CT was found to detect only a small minority of relapses [10] and was not found to be cost-effective [11]. Routine surveillance with positron emission tomographic (PET)/CT scans in first remission was found to have a low positive predictive value, high costs with no proven benefit [12]. For patients with early disease treated with radiation therapy alone, the method of relapse detection did not have a significant impact on the likelihood of successful additional treatment [13].

Our study had two objectives. First, to identify the proportion of recurrences noted by the patient as a result of new symptoms versus those detected via routine physical exams or tests ordered by the physician for asymptomatic patients. Second, to determine whether the method of detection of relapse is associated with significant differences in the time to diagnosis of relapse and postrelapse progression-free survival (PFS) and overall survival (OS).

**methods**

Following approval from the University of British Columbia—British Columbia Cancer Agency (BCCA) Research Ethics Board, we employed the BCCA Lymphoid Cancer Database, which prospectively collects comprehensive data on patients with a diagnosis of HL seen in British Columbia (BC) and contains information on 2531 patients with HL diagnosed between 1 April 1981 and 1 November 2011, to identify all patients with relapse of disease following a first confirmed CR of at least 3 months duration. We included patients aged between 15 and 65 years at the time of initial diagnosis. We excluded patients with primary refractory disease, and those with prior invasive cancers or positive serology for the human immunodeficiency virus. Of the 1917 patients (76%) with an initial CR, we focused on the 279 who relapsed >3 months later. After careful chart review, we excluded patients who had a different lymphoproliferative disorder at relapse (n = 6), refused or were ineligible for treatment (n = 4) or had ambiguous or missing information at relapse (n = 11).

After exclusions, we reviewed the remaining 258 cases to determine the proportion of recurrences detected as a result of new symptoms (PT group) versus routine preplanned physical exams or tests ordered by the physician in the absence of symptoms (MD group). We collected information regarding the nature of the presenting symptoms and the nature of the routine tests used by the physician to detect relapse. We then applied descriptive statistics to characterize and distinguish the PT and MD group. We used the Kaplan–Meier method to estimate survivals and log-rank comparisons to study the differences between the groups with regards to time to diagnosis of relapse and postrelapse PFS and OS.

Of relapse was defined as the time from original diagnosis of the HL to the time of notation by the patient (PT group) or physician (MD group) of the symptom or sign that led to detection of the relapse. Postrelapse PFS was the time from diagnosis of relapse to next relapse or death from any cause with patients free of relapse censored at the date of last follow-up. Postrelapse overall survival (OS) was the time from diagnosis of relapse to death from any cause with living patients censored at the date of last follow-up. Cox regression was used to adjust for potential confounders, including age at relapse (categorized as ≤50, 51–65, 66–75 or ≥76 years), sex and initial stage. All analyses were carried out using SAS for Windows version 9.2.

**results**

The characteristics at initial diagnosis of the patients in the PT and MD groups are shown in Table 1 and did not differ significantly. Of the total 258 relapses, 182 (71%, PT group) were brought to the attention of the physician by the patients themselves, most commonly because of recurrent lymphadenopathy (37% of all recurrences), B symptoms (19%), cough (8%) or other symptoms (7%). The remaining 76 relapses (29%, MD group) were detected in asymptomatic patients by the physician. Imaging accounted for the detection of most of these relapses (22% of all recurrences), mainly by chest X-ray and CT scan (9% and 7%, respectively). Physical exam accounted for the detection of 4% of all relapses. Laboratory tests, mostly complete blood counts and liver enzymes, led to the detection of 2% of all relapses.

The median time to diagnosis of recurrence was similar in both groups (PT group = 1.65 years; MD group = 1.95 years; P = 0.69). The postrelapse 5-years PFS did not differ significantly between the groups (PT group = 58.2%; MD group = 59.5%; P = 0.26) (Figure 1A), nor did the postrelapse 5-years OS (PT group = 78.3%; MD group = 86.0%; P = 0.40) (Figure 1B). Adjustment for age, gender and initial stage using a Cox proportional hazards model did not change these conclusions.

### Table 1. Characteristics at diagnosis of patients who relapsed >3 months after achieving CR with first-line treatment

<table>
<thead>
<tr>
<th></th>
<th>PT group</th>
<th>MD group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total, n (%)</td>
<td>182 (71)</td>
<td>76 (29)</td>
</tr>
<tr>
<td>Median age (years)</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>Male (%)</td>
<td>101 (55)</td>
<td>42 (55)</td>
</tr>
<tr>
<td>Disease stage</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I (%)</td>
<td>10 (6)</td>
<td>5 (7)</td>
</tr>
<tr>
<td>II (%)</td>
<td>82 (45)</td>
<td>30 (39)</td>
</tr>
<tr>
<td>III (%)</td>
<td>60 (33)</td>
<td>24 (32)</td>
</tr>
<tr>
<td>IV (%)</td>
<td>30 (16)</td>
<td>17 (22)</td>
</tr>
<tr>
<td>Decade of diagnosis of relapse</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1980s (%)</td>
<td>33 (18)</td>
<td>28 (37)</td>
</tr>
<tr>
<td>1990s (%)</td>
<td>73 (40)</td>
<td>20 (26)</td>
</tr>
<tr>
<td>2000s including 2011 (%)</td>
<td>76 (42)</td>
<td>28 (37)</td>
</tr>
<tr>
<td>Bulky (≥10 cm mass) (%)</td>
<td>55 (30)</td>
<td>15 (20)</td>
</tr>
<tr>
<td>IPPF score 0–3 (%)*</td>
<td>70 (38)</td>
<td>27 (36)</td>
</tr>
<tr>
<td>IPPF score 4 (%)†</td>
<td>17 (9)</td>
<td>5 (7)</td>
</tr>
</tbody>
</table>

*Data incomplete for calculation of IPPF on 133 patients (52%). PT, patient group; MD, physician group.
for time to diagnosis of recurrence ($P = 0.67$), postrelapse PFS ($P = 0.28$) or OS ($P = 0.22$). The types of treatments received at the time of relapse are outlined in Table 2.

discussion

We found that patients are much more likely to detect recurrence of HL than their physicians employing routine follow-up tests. More importantly, there was no difference in PFS or OS after relapse whether recurrence was detected by the patient or through routine screening by the physician. We could not demonstrate any benefit from early detection of HL recurrence with screening tests carried out on asymptomatic patients. In particular, we found that detection of relapse by the physician using routine preplanned tests did not confer any advantage in terms of subsequent PFS or OS. This strongly suggests that even more intensive, and therefore costly, surveillance is unlikely to prove advantageous. These findings are consistent with and validate the conclusions of previous smaller studies [9–13] with shorter follow-up. A study analogous to ours but examining second malignancies in testicular cancer survivors yielded similar results [14]. In light of the fact that most recurrences are detected by the patient, routinely reminding the patient of the importance of reporting any persistent new symptoms might further enhance detection of relapses, which can be done simply and at no cost.

The currently published guidelines for follow-up of patients in CR of HL are based on consensus and expert speculation rather than high-quality comparative data. They recommend routine imaging despite the lack of evidence that such follow-up improves the outcomes of recurrent HL. On the other hand, the disadvantages of routine imaging are numerous. First, exposure to ionizing radiation has well-documented potential deleterious health consequences [15, 16]. CT and PET scans in particular may pose a significant health hazard by exposing cured patients, many of whom have already received substantial exposure to such radiation during initial diagnosis and treatment, to excessive levels of ionizing radiation. The average CT scan of the chest, abdomen and pelvis exposes the patient to ~20 millisieverts (mSv) [15]. If NCCN guidelines, for instance, are strictly followed, the cumulative dose from follow-up alone over a period of 5 years may exceed 120 mSv. Recent evidence suggests that for every 10 mSv of ionizing radiation, there is a 3% increase in the risk of age- and gender-adjusted cancer over a mean follow-up period of 5 years (hazard ratio 1.003/mSv) [16]. Assuming a linear relationship between cumulative exposure and risk, the anticipated increased risk of cancer in patients who follow the NCCN guidelines could be as high as 1.36. The possibility that the risk of second malignancy may be this high emphasizes the importance of eliminating unnecessary diagnostic testing.

Other disadvantages associated with routine screening are cost and induction of anxiety. CT is cost-ineffective for the follow-up of patients with HL after initial treatment [11]. In addition, the work-up necessary to investigate false-positives adds not only to the cost, but also to the anxiety of HL survivors. Routine surveillance scans have been found to exacerbate anxiety in long-term lymphoma survivors [17, 18].

It is reasonable to ask why surveillance testing appears to be ineffective. This reflects the low and rapidly diminishing prior probability of relapse in these patients, which, in turn, reflects the frequency with which patients with HL are cured by primary treatment. In addition, because relapse occurs infrequently there is a high probability of encountering false-positive test results, which, in turn, are likely lead to additional potentially costly and invasive testing. For example, suppose a test is carried out on an asymptomatic patient twice during the period from 18 to 30 months from initial diagnosis. This is approximately the second year of follow-up of treatment and is a fixed interval during which the likelihood that a HL patient will relapse is <5%. If such a test, perhaps whole body CT

Table 2. Type of treatment received at the time of relapse

<table>
<thead>
<tr>
<th>Treatment</th>
<th>PT group</th>
<th>MD group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Too frail for potentially curative treatment (%)</td>
<td>21 (12)</td>
<td>8 (11)</td>
</tr>
<tr>
<td>RT alone (%)</td>
<td>11 (6)</td>
<td>6 (8)</td>
</tr>
<tr>
<td>Standard chemotherapy ± RT (%)</td>
<td>38 (21)</td>
<td>21 (28)</td>
</tr>
<tr>
<td>High-dose chemotherapy with stem cell support ± RT (%)</td>
<td>110 (60)</td>
<td>40 (53)</td>
</tr>
</tbody>
</table>

RT, radiotherapy; PT, patient group; MD, physician group.
scanning or PET scanning, has a 5% risk of false positivity, which is probably a low estimate for such complex tests, for each test carried out the probability of being truly positive is only 1 of 20 (5%) and being falsely positive 2 of 20 (10%). In this sense, harm is twice as likely as gain. Greater frequency of testing, such as every 3 months as suggested by some authorities, or later application of the testing, for example during the fourth year of follow-up during which the risk of relapse has fallen to <2%–3% for that year, only exacerbates the risk of false positivity and cost-ineffectiveness, because the likelihood of relapse diminishes but the probability of false positivity remains constant.

Our study has several potential limitations. First, because this experience reflects a mix of academic and community hematology–oncology practices and spans 30 years of experience the type and frequency of follow-up exams could not be rigidly required of physicians and patients nor could we accurately capture the frequency of their performance in our database. However, given that the baseline characteristics of the two groups were similar (Table 1), in particular with regards to prognostic indices, there is no reason to believe that the intensity of follow-up differed between them. In addition, any bias toward a greater tumor burden and therefore worse prognosis should have disproportionately adversely affected the patients found to have recurrence due to self-detected symptoms, but these patients did just as well as those whose relapse was identified by the physician. Finally, we did not uncover any evidence that the ratio of patient- to physician-discovered relapse changed over the three decades of our study despite the ever increasing availability and sensitivity of imaging equipment (Table 1). A second limitation reflects the 30-year period that our database spans during which the types of tests available for follow-up have changed (Table 1). To minimize this potential source of bias, we only included patients managed since 1981 when CT scanning became widely available in BC and follow-up guidelines became uniformly available. Also, of note, reflecting our exclusion criteria, our findings apply only to patients who relapsed after entering a first CR. Patients who achieve true partial responses and have primary refractory disease may benefit from a different approach to follow-up. Thirdly, the data collected reflect the experience within a single Canadian province where health care is publicly financed and universally available. This may theoretically limit its applicability to other settings where availability of testing may be dependent on ability to pay; however, it seems most likely that the cost implications in such systems would amplify the value of avoiding unproductive testing. Thus, it appears unlikely that relapse detection would follow a more favorable pattern in healthcare systems where there are financial obstacles to obtaining diagnostic tests. Finally, our study has not examined the role of routine follow-up testing for detection of long-term complications of therapy such as secondary malignancies, cardiovascular diseases or endocrine disorders.

In conclusion, our study does not demonstrate any benefit, in the absence of symptoms or suggestive physical examination findings, to performing laboratory testing or imaging of HL patients who have maintained first CR for at least 3 months. Given the significant risks and potential toxicity of such follow-up testing, in particular exposure to high cumulative doses of ionizing radiation and induction of anxiety, current guidelines are difficult to justify and should be reviewed. In addition, prospective trials should also be considered to help define the optimal follow-up of HL survivors.

acknowledgements

The authors thank Jane Donaldson and Suman Singh for data collection and database maintenance and the physicians and patients of British Columbia for inclusion of information they provided.

funding

Supported in part by the Turner Family Lymphoma Outcomes Fund and the Mary Toye Memorial Fund [no grant numbers].

disclosure

The authors have declared no conflicts of interest.

references


NIH-defined graft-versus-host disease and evidence for a potent graft-versus-leukemia effect in patients with acute lymphoblastic leukemia

T. H. Terwey1*, T. M. Le Duc1, P. G. Hemmati1, P. le Coutre1, M. Nagy2, P. Martus3, B. Dörken1 & R. Arnold1

1Department of Hematology, Oncology and Tumor Immunology; 2Institute of Legal Medicine and Forensic Sciences, Charité – University Medicine Berlin, Berlin; 3Department of Medical Biometry, Friedrich Karl University Tübingen, Tübingen, Germany

Background: The prognostic value of the NIH consensus criteria for graft-versus-host disease (GVHD) is not well defined yet.

Patients and methods: We analyzed NIH-defined GVHD in 147 acute lymphoblastic leukemia (ALL) patients.

Results: The cumulative incidence of classic acute GVHD (aGVHD), late aGVHD and chronic GVHD (cGVHD) was 63%, 12% and 41%, respectively. cGVHD was subclassified as classic versus overlap syndrome in 40% versus 60% of cases. In multivariate Cox regression analysis with GVHD as time-dependent covariate, classic aGVHD grade III/IV had a negative impact on overall survival (OS) due to higher non-relapse mortality. cGVHD of any grade was associated with superior OS, which was due to lower relapse incidence. Classic cGVHD versus overlap syndrome had no differential impact. In 44 patients without GVHD after transplant who received donor lymphocyte infusions (DLI), the cumulative incidence of classic aGVHD, late aGVHD or cGVHD was 60%, 5% and 57%. Occurrence of cGVHD after DLI was associated with improved OS due to lower relapse incidence.

Conclusions: The NIH consensus criteria for GVHD clearly define prognostic subgroups in patients transplanted for ALL. The improved OS in patients developing cGVHD after transplant or DLI gives clear evidence for a potent graft-versus-leukemia effect in this indication.

Key words: acute lymphoblastic leukemia, GVHD, GVL, NIH consensus criteria

© The Author 2012. Published by Oxford University Press on behalf of the European Society for Medical Oncology. All rights reserved. For permissions, please email: journals.permissions@oup.com.