Identifying subgroups among poor prognosis patients with nonseminomatous germ cell cancer by tree modelling: a validation study

M. R. van Dijk 1*, E. W. Steyerberg 1, S. P. Stenning 2 & J. D. F. Habbema 1

1Department of Public Health, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands; 2Medical Research Council, Clinical Trials Unit, London, UK

Received 19 January 2004; revised 22 April 2004; accepted 23 April 2004

Background: In order to target intensive treatment strategies for poor prognosis patients with nonseminomatous germ cell cancer, those with the poorest prognosis should be identified. These patients might profit most from more intensive treatment strategies. For this purpose, a regression tree was previously developed on 332 patients. We aimed to evaluate the performance and structure of this tree.

Patients and methods: The previously developed tree was applied to 456 patients with a poor prognosis as defined by the International Germ Cell Cancer Collaborative Group (IGCCCG). Next, we developed a new tree to evaluate whether a similar structure to the previous tree was found. We assessed the internal validity of the new tree, and compared the 2-year survival estimates of each subgroup together with the discriminative ability for both the previously developed and the new tree. Discriminative ability was measured by a concordance (c) statistic, which varies between 0.5 (no discrimination) and 1.0 (perfect discrimination).

Results: The 2-year survival estimates in the IGCCCG data ranged from 33% to 63%. The ordering of the subgroups was different and discriminative ability was lower than originally found (c = 0.56 in the IGCCCG data versus 0.63 originally). The new tree differed considerably from the original tree, and identified poor prognosis subgroups with 2-year survival estimates from 38% to 73%. Internal validation showed similar discriminative ability for the new tree and the original tree (c = 0.59 versus 0.56).

Conclusions: The previously developed tree showed poor validity with respect to discriminative ability and the stability of its structure. The performance of the new tree was also unsatisfactory. Given the low proportion of patients categorised as poor prognosis, it seems that the potential to identify further subgroups with the currently available patient characteristics is limited.

Key words: germ cell cancer, poor prognosis, tree modelling, validation

Introduction

Patients with metastatic nonseminomatosus germ cell tumours nowadays have a long-term cure rate of >80%, because of the ability of cisplatin-based chemotherapy to cure advanced disease [1–4]. Because of the high overall cure rate, interest has shifted to reducing treatment-related toxicity for patients with a good prognosis [5]. On the other hand poor prognosis patients should be considered for more aggressive treatment strategies [6–8]. The International Germ Cell Cancer Collaborative Group (IGCCCG) developed the International Germ Cell Consensus (IGCC) classification to distinguish patients according to prognosis [9]. The IGCC classification is currently widely applied. It distinguishes patients with good, intermediate and poor prognosis. The poor prognosis group consists of ~15% of all patients and is characterised by the presence of mediastinal primary site, non-pulmonary visceral metastases or poor tumour marker levels. Long-term survival following standard treatment may be ~ 50% [9]. To further improve the long-term survival in poor prognosis patients, those who are most likely to fail standard treatment should be identified. These patients are most likely to profit from novel chemotherapy approaches, such as dose intensification and high-dose chemotherapy with stem-cell support [10]. Dose intensification with either sequential or alternating non-cross-resistant chemotherapy has shown promising results in...
non-randomised trials [11], but this has not been confirmed in randomised clinical trials [10, 11].

Studies conducted by the Memorial-Sloan Kettering Cancer Center [12, 13] and the German Testicular Cancer Group [7, 14] showed beneficial effects for high-dose chemotherapy with stem-cell support and high-dose chemotherapy, respectively. However these results were based on non-randomised trials and comparisons with historical controls. To confirm these results, two randomised, multicentre trials are currently being conducted in the USA and Europe [15]. Furthermore, the identification of subgroups among poor prognosis patients would allow for a more accurate estimate of the individual patients’ chances of survival, and increase the comparability of results from clinical trials [16].

Kollmannsberger et al. [16] used tree modelling as an explorative method to identify important risk factors within a group of poor prognosis patients, as defined by the IGCC classification, and to find subsets of patients differing in prognosis. They developed a regression tree based on data of 332 poor prognosis patients as defined by the IGCC classification (Kollmannsberger tree). The risk factors visceral metastases, primary site and abdominal mass were used. This resulted in a tree with five poor prognosis subgroups (Figure 1A). The subgroups differed in 2-year survival, ranging from 49% to 84% [16].

Tree models are attractive ways to identify subsets of patients because they are easy to apply and interpret. They have few restrictions, which makes them suitable for finding interactions between risk factors [17–20]. Furthermore, trees have more resemblance with the way clinicians make decisions than linear models [17]. On the other hand, this flexibility makes trees ‘data hungry’. Use of relatively small datasets will lead to unstable tree models, and optimism in the performance of the model due to overfitting [17, 21, 22]. Kollmannsberger et al. [16] recognise these problems and the limitations of their tree. Some subgroups only had a small number of patients, and their identification may be the result of pure chance. Such subgroups may not be present when new data are considered. Furthermore, survival estimates of small groups are often unreliable. This was illustrated by the group of patients with visceral metastases and primary site testis, in which patients with an abdominal mass had a higher 2-year survival [72%; 95% confidence interval (CI) 64–80%] than patients without (52%; 95% CI 27–77%). Kollmannsberger et al. [16] therefore propose that further confirmatory studies in other poor prognosis patients cohorts are needed, before the tree can be used in practice.

The aim of the present study was to evaluate the validity of the Kollmannsberger tree. We consider two aspects of validity: performance and the structure of the tree model. We evaluated the performance of the Kollmannsberger tree by applying the tree to the poor prognosis patients in the IGCCCG database [9]. Furthermore, we developed a new tree for the poor prognosis patients in the IGCCCG database to study whether its structure, that is the selection and hierarchy of risk factors, was similar to the Kollmannsberger tree.

**Figure 1.** Trees for poor prognosis patients with nonseminomatous germ cell cancer with 2-year survival, 95% confidence interval and number of patients (n). (A) Kollmannsberger tree applied to the Kollmannsberger data. (B) Kollmannsberger tree applied to the International Germ Cell Cancer Collaborative Group (IGCCCG) poor prognosis data. (C) International Germ Cell Consensus (IGCC) tree applied to the IGCCCG poor prognosis data.

**Patients and methods**

**Patients**

The Kollmannsberger tree was based on data from 332 patients with metastatic nonseminomatous germ cell tumours, from prospective clinical trials conducted between 1984 and 1997 [23, 24]. All patients were treated for poor prognosis disease, as defined by the IGCC classification [9], with either cisplatin–etoposide–ifosfamide (PEI) or cisplatin–etoposide–bleocycin (PEB).
To validate the Kollmannsberger tree, we used the 495 poor prognosis patients with nonseminomatous germ cell cancer from the IGCCCG database, which consists of 5202 adult male patients. Poor prognosis was defined by the presence of any of the poor risk factors mediastinal primary site, (non-pulmonary) visceral metastases, α-fetoprotein (AFP) poor (>10000 ng/ml), human chorionic gonadotrophin (HCG) poor (>10000 ng/ml) and lactate dehydrogenase (LDH) poor (>10 times the upper limit of normal) [9]. Patients were treated between 1975 and 1990 with cisplatin-based chemotherapy. Data were collected on age, primary site, date of diagnosis, levels of serum AFP, HCG and LDH, nodal disease in the abdomen, mediastinum, neck, lung metastases, spread to other visceral sites such as liver, bone and brain, and on treatment details such as previous therapy [9]. Data for 39 patients were excluded because of missing values on the risk factors age, and lung, liver, bone and brain metastases. The endpoint was overall survival, calculated from the beginning of chemotherapy.

Statistical analyses
We assessed the performance of the Kollmannsberger tree by applying it to the 456 poor prognosis patients in the IGCCCG database. Two-year overall survival was calculated using the Kaplan–Meier method.

IGCC tree
Furthermore, a new tree was developed in the poor prognosis patients in the IGCCCG database and the result compared with the Kollmannsberger tree to evaluate its structure, that is the selection and hierarchy of the risk factors. We will refer to this new tree as the IGCC tree. For the development of this tree we used the same candidate risk factors and coding as Kollmannsberger et al. [16]. We first determined the risk factor that best split the data into two subgroups, leading to the largest decrease in prediction error. Splitting continued until a group reached a minimum size of five, or until no further improvement in discrimination could be made, based on the loss in exponential log-likelihood. The full tree, which might be too complex and overfit, was pruned using 10-fold cross-validation. All trees within one standard error of the lowest cross-validated prediction error were considered as equivalent. A final tree was selected from these equivalent trees [19, 20, 25].

Modelling was performed with S-plus version 2000 using the RPART library, which contains a recursive partitioning method for survival data. The RPART library (rpart2.zip) and manual (rpart2doc.zip) can be found at http://www.stats.ox.ac.uk/pub/SWin.

For comparison we fitted a Cox regression model using the same risk factors in the poor prognosis patients in the IGCCCG database. All risk factors were entered in the model and the final model was obtained with a backward stepwise selection procedure using a P-value of 0.05.

Predictive performance
We determined the discriminative ability, to indicate the predictive performance of the Kollmannsberger tree, the IGCC tree and the Cox regression model. The discriminative ability indicates how well a model can distinguish between patients with different survival expectations and was measured by a concordance statistic (c-statistic). For binary outcomes, the c-statistic is identical to the area under the ROC curve [26, 27]. The c-statistic for survival data estimates the probability that for a randomly chosen pair of patients, the one having the higher predicted survival is the one who survives longer [26]. The C-statistic varies between 0.5 and 1.0 for sensible models. A predictive model with a c of 0.5 has no predictive value, while a model with a c of 1.0 discriminates perfectly between patients differing in survival. c-statistics were computed for the Kollmannsberger tree, when applied to the Kollmannsberger patients or the IGCCCG patients, for the IGCC tree and for the Cox regression model.

The steps taken in the development of the IGCC tree were internally validated by taking 100 random bootstrap samples. The development process of the tree was repeated for every bootstrap sample and the resulting tree tested on the original sample, to estimate and correct for the optimism in discriminative ability [28, 29]. The original sample hence served as the test sample for models developed in the bootstrap samples. The Cox regression model was validated according to the same procedure. The standard error of the corrected c-statistic was taken from the empirical distribution of the c-statistics in the test sample. This standard error was used to calculate the 95% CI of the optimism-corrected c-statistic.

Results
Patient characteristics are given in Table 1, both for the 456 poor prognosis patients of the IGCCCG study and the 332 patients of the Kollmannsberger study. More than half of the IGCCCG poor prognosis patients had primary site testis (67%), lung metastases (62%) or abdominal masses (70%). Sixty-three per cent of the patients had poor AFP, HCG or LDH levels. The presence of liver metastases was common (34%). The distribution of the patient characteristics age, lung metastases, visceral metastases, abdominal masses, number of metastatic sites and tumour markers, combined as well as separate, and follow-up time was largely similar for the IGCCCG and the Kollmannsberger studies. Disease progression occurred in 252 of the IGCCCG poor prognosis patients. Of the 223 patients who died, 213 were categorised as disease-related. The corresponding 2-year survival was 56% (95% CI 52–61%) for the IGCCCG poor prognosis patients. This differs from the Kollmannsberger data, where 2-year survival was 72% (95% CI 67–77%).

The Kollmannsberger tree was applied to the poor prognosis patients in the IGCCCG database, as presented in Figure 1B. The split according to the presence of visceral metastases resulted in two subgroups with only slightly different 2-year survival (59% and 53%).

Both branches were split further according to primary site, resulting in two final groups in the no visceral metastases branch, with 2-year survival of 63% and 50%, and in the visceral metastases branch in one final group with 2-year survival of 33% and one further subgroup with 2-year survival of 56%; the last final groups from this branch were defined by the presence of abdominal mass, with similar 2-year survival (56% and 53%). As can be seen by comparing Figure 1B and A, respectively, the 2-year survival estimates for the five final groups identified in the poor prognosis patients in the IGCCCG database by the Kollmannsberger tree were less extreme than in the original Kollmannsberger tree. This was reflected by the lower discriminative ability in the IGCCCG poor prognosis data (c = 0.56; 95% CI 0.49–0.64) compared with the original data (c = 0.63; 95% CI 0.56–0.70).

The newly developed IGCC tree is presented in Figure 1C. Trees with two to six groups gave equivalent performance based on the loss in exponential log-likelihood. However, a tree with five final groups was chosen for fair comparability with the Kollmannsberger tree. The 2-year survival ranged
from 38% to 73%. The principal determinant of survival was the total number of metastases, where three or fewer metastases resulted in a subgroup with a 2-year survival of 61% and the presence of more than three metastases in a final group with a 2-year survival of 38%. The next split was made by primary site, resulting in a subgroup of patients with testis as primary site and a 2-year survival of 64% and a final group of patients with mediastinal primary site and a 2-year survival of 49%. A further distinction was made by the size of an abdominal mass, resulting in a 2-year survival of 73% for patients with a mass \( \leq 10 \text{ cm} \), and 56% for a mass \( >10 \text{ cm} \). Finally, the presence of lung metastases resulted in two final groups with a 2-year survival of 67% for patients without lung metastases and 42% for patients with lung metastases.

Thus, both the Kollmannsberger tree and the new IGCC tree selected primary site and abdominal mass as important risk factors, although the Kollmannsberger tree selected the presence of abdominal mass rather than size. Furthermore, the presence of lung metastases was used in the IGCC tree, but not in the Kollmannsberger tree. The apparent discriminative ability of the new IGCC tree was similar to the discriminative ability of the Kollmannsberger tree, with a c-statistic of 0.63 (95% CI 0.62–0.72). Internal validation revealed an optimism in c-statistic of 0.04, leading to an optimism-corrected estimate of \( c = 0.59 \) (95% CI 0.54–0.63) for the new IGCC tree when applied to future patients similar to those included in the IGCCCG database.

The Cox regression model selected the risk factors primary site, presence of abdominal mass, size of abdominal mass, total number of metastases, AFP and tumour markers combined. The discriminative ability of the Cox regression model was slightly higher, with a c-statistic of 0.64, but decreased to 0.61 after correction for optimism.

**Discussion**

The previously developed tree to identify subgroups among poor prognosis patients with nonseminomatous germ cell cancer showed poor validity in the poor prognosis patients in the IGCCCG database. First, the discriminative ability of
the Kollmannsberger tree was substantially lower (c = 0.56) than in the original data (c = 0.63). Secondly, a new tree was developed in the poor prognosis patients in the IGCCCG database, which used the number of metastases, primary site, size of abdominal mass and the presence of lung metastases to identify subgroups, whereas the Kollmannsberger tree used the presence of visceral metastases, primary site and abdominal mass as risk factors. The discriminative ability of the new tree, c = 0.59 at internal validation, was similar to the Kollmannsberger tree (c = 0.56).

In our case, the selected risk factors were rather similar in the trees developed with the Kollmannsberger data and the IGCCCG poor prognosis data (presence versus size of abdominal mass and presence versus number of metastases). The structure of the trees, however, was very different. Since primary site was an important risk factor in patients with and without visceral metastases in the Kollmannsberger tree, both risk factors can be interpreted statistically as main effects. In the IGCC tree, primary site occurred only once, and no main effects were present except for the number of metastases. Furthermore, the trees fitted in the bootstrap samples varied in size, the smallest tree having only two groups and the largest tree 20. A tree size of four was most prevalent (14% of cases). The flexibility in structure and tree size led to optimism in the performance of the tree developed on the IGCCCG poor prognosis data, where the c-statistic was expected to decrease from 0.63 to 0.59 according to a bootstrap validation technique. Hence, the flexibility of tree modelling may have more cons than pros in small data sets, that is datasets with relatively few deaths. Larger datasets are required for reliable application of tree modelling.

The discriminative ability of a Cox regression model was slightly higher than the regression tree models with a c-statistic of 0.61 at internal validation. A Cox regression model was not available for the Kollmannsberger data. Should such a model be applied to the Kollmannsberger data, the selection and the number of risk factors might differ. However, the structure is likely to be similar for a given selection of risk factors, since we usually fit main effects only in a regression model. In relatively small datasets, this may be the most sensible, since we have insufficient statistical power to identify important interaction terms [30].

To assess whether the Kollmannsberger tree can be generalised to other patients, we applied it to poor prognosis patients from the IGCCCG dataset. The comparison of the Kollmannsberger and the IGCC tree was, however, limited by the differences in the two datasets. Although the distribution of risk factors in the Kollmannsberger and IGCCCG poor prognosis data was largely similar, there was a difference in 2-year survival (72% and 56%, respectively). These differences may reflect the different time periods in which the data were collected (Kollmannsberger 1984–1997, IGCCCG 1975–1990). Owing to improved treatment strategies, survival has increased over time [4, 31]. Patients in the Kollmannsberger study were treated with regimens of either PEB or PEI. Although all poor prognosis patients in the IGCCCG database were treated with cisplatin-based chemotherapy, patients from the late 1970s and early 1980s were probably treated with the cisplatin–vinblastine–bleomycin regimen rather than PEB. The differences between the populations suggest that a more recent population of poor prognosis patients might be more suitable to assess the generalisability of the Kollmannsberger tree. Ideally, to assess the differences between the Kollmannsberger and the IGCCCG poor prognosis data and to make an honest comparison between both trees, the new IGCC tree should be applied to the Kollmannsberger data.

Better performance might be achieved by adding stronger risk factors that have not been used before in the classification of patients with germ cell cancer. Besides pre-treatment characteristics, the rate of tumour marker decline during the first two cycles of chemotherapy has been identified as an important risk factor. Rate of tumour marker decline predicted outcome in 189 patients, independent of risk status as defined by the IGCC classification, especially in poor prognosis patients [33]. Similar results were found in 139 poor prognosis patients in a previously conducted study [32]. These results will be validated with data from a multicentre, randomised clinical trial [33]. Furthermore, promising research is being carried out on the prognostic value of molecular and genetic markers. Knowledge on the role of such markers will not only allow for a better understanding of the development and progression of testicular germ cell cancer, but may also lead to a more refined assessment of prognosis and better management of germ cell tumours [34, 35].

In conclusion, survival of IGCC poor risk patients in the present day may have improved compared with the historical IGCCCG data. This justifies the investigation of poorer risk subgroups, although the difficulties in evaluating new treatment approaches through randomised trials in these small groups must be acknowledged. The performance of the current regression trees was unsatisfactory. The currently available risk factors are not strong enough to clearly identify subgroups among poor prognosis patients with nonseminomatous germ cell cancer, who comprise a small subgroup (~15%) of metastatic germ cell tumour patients. A new model might incorporate molecular and genetic markers in addition to the risk factors currently incorporated in the IGCC classification. We suggest the use of Cox regression for the construction of such a new model, rather than tree modelling. This method is proven to be more stable and gives less optimistic results, especially in smaller datasets. Tree modelling can give insight into possible interactions between risk factors provided sufficient data is available, but should be restricted to exploratory analyses.

Acknowledgements

We would like to thank the members of the IGCCCG for kindly providing their data for the analyses in this paper. Members of the IGCCCG and participating centres have been listed previously [9]. This work was supported by the Netherlands