Predicting the duration of chemotherapy induced neutropenia: new scores and validation

S. San Matías¹*, M. Clemente¹, V. Giner-Bosch¹ & V. Giner²

¹Departamento de Estadística e Investigación Operativa Aplicadas y Calidad, Universidad Politécnica de Valencia, Valencia, Spain.  
²Oncology Unit, Hospital de Sagunto, Sagunto, Spain.

*Correspondence to: Dr. S. San Matías, Departamento de Estadística e Investigación Operativa Aplicadas y Calidad, Universidad Politécnica de Valencia, Camino de Vera s/n, 46022 Valencia, Spain. Tel: +34 963877490; Fax: +34 963877499; E-mail: ssanmat@eio.upv.es

Background: The objective of this work was to develop predictive models to classify febrile neutropenic patients into two groups, according to the prediction of the duration of the chemotherapy-induced neutropenia (CIN) episode.

Patients and methods: For this analysis, 148 patients with solid tumours and an episode of febrile neutropenia (FN) were eligible. A score was attributed to each drug given in a chemotherapy treatment, according to its expected toxicity. Two new scores were proposed basing only on this classification. The first one is a combination of the individual scores and the second one was built using statistical techniques as cluster analysis and classification trees.

Results: Statistical techniques produced the best score, that was able to discriminate two groups of patients with statistically different neutropenia durations, with median durations until hematological recovery of ANC $\geq 2 \times 10^9/l$ being 3 versus 2 days ($P = 0.002$).

Conclusions: Using the aggressiveness of the cytotoxic agents in a chemotherapy regimen, we offer new improved alternatives to previous studies in order to identify which patients will need longer times to recover from neutropenia. Our proposal discriminated two groups of patients, one of them needing 50% more of time to recover than the other.

Key words: chemotherapy, classification trees, cluster, duration of neutropenia, febrile neutropenia, statistical techniques.

Background

The neutropenia is a disorder characterized by an abnormally low number of neutrophils in the blood. Particularly, chemotherapy-induced neutropenia (CIN) is the most common side effect associated to the administration of anticancer drugs. Up to 25% of patients treated with chemotherapy are likely to develop a FN episode (see for example Crawford et al. [1]),
although this percentage could increase up to 96% in some particular type of tumors (Crawford et al. [2]). Previous studies relating to the apparition of neutropenia after administration of chemotherapy can be found in Blay et al. [3], Silber et al. [4], Ray-Coquard et al. [5], Kondo et al. [6] and Wilson-Royalty et al. [7].

The Common Toxicity Criteria of the National Cancer Institute established a scale of four grades for neutropenia, according to the absolute neutrophil count (ANC): grade 1, ANC ≥1.5 to <2×10⁹/l; grade 2, ANC ≥1.0 to <1.5×10⁹/l; grade 3, ANC ≥0.5 to <1.0×10⁹/l; grade 4, ANC <0.5×10⁹/l (see [1]). A patient with FN will be very susceptible to suffer life-threatening complications including death, and this is related to the duration and severity of the FN episode [8]. Moreover, the higher duration of neutropenia, the higher infection risk, so it is fundamental to determine the CIN duration at the onset of a febrile neutropenic episode.

It is now accepted that it exists a relation between the aggressiveness of the chemotherapy regimen and the neutropenia duration (see [3]). Recently, Lalami et al. [9] have developed a model which aims to predict the duration of chemotherapy induced neutropenia according to the aggressiveness of the cytotoxic regimen. They assign a score to each chemotherapy regimen and classify patients into two groups according to their values of such score. In their work, they show that the medians duration of CIN in these two groups are significantly different. However, Giner et al. (2007, personal communication) applied the regimen based scores proposed in [9] to a new sample from a single institution without succeeding in the estimation of the duration of CIN episodes.

In this work, our objective has been to design a predictive system for the neutropenia duration, alternative to the model by Lalami et al. in [9]. We propose new different methodologies to classify patients into two groups of expected low and high duration. As in [9], we also base our proposals only on the hematological toxicity of each chemotherapy regimen.

**Patients and methods**

Our sample was obtained from patients of the Oncology Unit of the Hospital of Sagunto (Valencia, Spain). The data were collected from June 1997 to October 2006. All our patients were outpatients, that is, they had not been admitted to the hospital at the onset of FN. The eligibility criteria for the patient selection were as follows:

1. Diagnosis of malignancy treated with chemotherapy. All the patients suffer solid tumors.
2. Neutropenia caused by this treatment. ANC (Absolute Neutrophil Count) < 0,5×10⁹/l
3. Temperature > 38 ºC (Febrile Neutropenia)
4. Age ≥ 16 years
5. Treatment with an appropriate initial empiric antibiotic regimen
6. First episode for each patient
We focused our study on the time to recovery for an ANC $\geq 2 \times 10^9/l$ due to the fact that in our sample the time to recovery is in general very short. We will denote as $d1$ the time (days) to overcome grade 1 of CIN (from the beginning of the episode). For the sake of comparison, we assumed some of the premises in [9]. Namely:

1. The beginning of neutropenia was settled as the first day with the documentation of grade 4 of neutropenia.
2. Our models aim to predict the duration of CIN according only to the aggressiveness of the chemotherapy regimen.
3. The aggressiveness of the chemotherapy regimen is taken into account by giving to each individual drug a score (ranging from 0 to 4), according to its expected hematological toxicity basing on data from the literature. The assignment of the chemotherapy score to each drug used in our sample was:
   - Score 0: bleomycin, leucovorin, cetuximab, estramustine
   - Score 1: 5-fluorouracil, cisplatin, fludarabin, procarbazine, rituximab
   - Score 2: gemcitabine, melphalan, methotrexate, mitoxantrone, mitomycin C, raltitrexed, vinblastine, vinorelbine, vincristine
   - Score 3: carboplatin, cyclophosphamide, epirubicin, ifosfamide, oxaliplatin, adriamycin
   - Score 4: irinotecan, docetaxel, paclitaxel, topotecan, etoposide

Among the different proposals of Lalami et al. in [9], the best performance was achieved by the aggregated score $S2$, that they obtained by summing up the individual scores of each drug into the chemotherapy regimen. Then, a threshold of 8 points is used in order to classify the regimen either into a group of low expected duration of neutropenia (level 1) or high expected duration (level 2).

Next we present our two new proposals which aim also to obtain a global classification score for a particular chemotherapy regimen.

**New score R1**

With the same idea of $S2$ in [9], our score $R1$ combines the individual scores of each drug included in a chemotherapy regimen. However, our proposal differs in the aggregation type, because we think that the addition could not be the best way of combining the individual scores anymore if the different drugs supplied to a patient would interact somehow. We have studied two new ways of combination:

- Magnifying trend ($M$-scores). They consist on increasing weighted sums. Our results showed that the $M$-scores were not discriminative and so we will not go into this trend in depth.
- Reducing trend ($R$-scores). They consist on decreasing weighted sums. The underlying hypothesis is that some noxious effects may be common when
administrating drugs with similar components, and then they could produce lower toxicity than it was expected. The best performance of our reducing trend scores was achieved by the so-called score R1.

Let us consider for each patient the cytotoxic drugs supplied and their individual scores according to the expected hematological toxicity. Then, in order to compute the value of R1 for each patient it must be applied Algorithm 1, that we present now.

**Algorithm 1 – Score R1**

Step 1. For each patient, do the following:

1. Order the scores in ascending way (0 to 4).
2. Assign descending weights to each score, starting by 1, 0.9, 0.8, etc. Add the scores multiplied by the corresponding weights and assign to R1.
3. If R1 < 6, then classify the patient into the low duration group. If R1 ≥ 6, then classify the patient into the high duration group.

Step 2. Compute the median duration of patients in both duration groups.

Step 3. Assign to each episode the median duration according to the membership duration group.

In our patients, R1 ranged from 0.9 to 8.9. Its median was 5.6, while 75% of patients had a value of R1 lower than 6.6.

**New score T1**

First of all, we classified our sample of patients into two groups: the group of lowest durations (cluster Low), and the group of highest durations (cluster High). The threshold to carry out this classification was determined by a cluster analysis (see for example Mardia [10]), performed with the variable $d1$. This procedure gave us as a result two groups of patients as separated as possible with respect to the neutropenia duration. The first group or cluster Low was formed by the 77 patients with lowest durations, whose maximal duration was equal to 3 days. The 29 patients with higher durations formed the so-called cluster High, with minimal duration equal to 4 days. Now, let us define for each patient:

$$Gi = \text{number of different drugs with score } i \text{ supplied to a patient into the chemotherapy, for } i=0 \text{ to } 4.$$

Our goal has been to predict membership of patients in cluster Low or High from their measurements on variables $G$, using statistical methodology. Particularly, we have used the statistical technique called CART (Classification and Regression Tree, see Breiman et al. [11]).

A CART analysis is a form of binary recursive partitioning. Its purpose is to determine a set of if-then logical (split) conditions that permit accurate prediction or classification of
cases. The tree is structured as a sequence of simple questions based on the predictor variables $G$, and the answers to these questions trace a path down the tree.

Each group of patients is represented by a “node” in the classification tree. The root node is the original sample, which has a particular proportion of patients belonging to cluster $Low$ (72.6%). This node can be splitted into two smaller groups or child nodes, using a question about the value of one of the variables $G$. CART searches for questions that split nodes in such a way that the proportion of cluster $Low$ in the two child nodes is as different as possible. This binary partitioning process can be applied over and over again. In Figure 1 we show the classification tree obtained with the non-censored cases of our sample.

**Figure 1.** Classification tree to predict membership in low/high duration group.
When the tree is completed, each of the final nodes can be seen as the collection of all the chemotherapy regimens having the number of cytotoxic drugs that defined the node. Now, for all these chemotherapy regimens we assign the value of $T_1$ as the proportion of patients of cluster \textit{High} in this node. If a chemotherapy regimen has a value of $T_1$ less than 26%, we will predict LOW duration of neutropenia. In other case, we will predict HIGH duration. Algorithm 2 summarizes the computation and use of $T_1$.

\textit{Algorithm 2 – Score $T_1$}

\textbf{Step 1.} For each patient, do the following

- \textbf{Step 1.1.} If $G_1=G_2=0$ and $G_3\geq3$ and $G_4\geq1$, then $T_1=50$
- \textbf{Step 1.2.} If $G_1\geq1$ and $G_2=0$, then $T_1=32$
- \textbf{Step 1.3.} If $G_1=G_4=0$ and $G_2\geq1$, then $T_1=27$
- \textbf{Step 1.4.} If $G_1=G_2=0$ and $1\leq G_3\leq2$ and $G_4\geq1$, then $T_1=25$
- \textbf{Step 1.5.} If $G_1\geq1$ and $G_2\geq1$ and $G_4=0$, then $T_1=12$
- \textbf{Step 1.6.} Otherwise, then $T_1=0$
- \textbf{Step 1.7.} If $T_1 < 26$, then classify the patient into the low duration group.
  
  \hspace{1cm} If $T_1 \geq 26$, then classify the patient into the high duration group.

\textbf{Step 2.} Compute the median duration of patients in both duration groups.

\textbf{Step 3.} Assign to each episode the median duration according to the membership duration group.

The median of $T_1$ in our sample was 27, ranging from 0 to 50. Only 25% of patients had a value of $T_1$ greater than 32.

\textbf{Results}

We present now some descriptive analysis of our sample. One hundred and forty eight patients were eligible for the survey but 42 were excluded because the lack of data at the onset of FN. In the following, we will consider only the 106 non excluded cases, 21 of whom had a subsequent episode of FN that was not considered for the study. There were 49 women (46.2 \%) and 57 men (53.8\%), and their ages ranged from 16 to 84, being the average age around 62 years and the median age 65.

In our sample we found 19 different types of solid tumours. However, 3 of them represent the 70\% of the total cases: breast (31 cases, 29.2\%), lung (28 cases, 26.4\%) and head-neck (17 cases, 16\%). Other tumours represent 28.4\% of the total number of cases. With respect to the number of different drugs supplied to the patients into the chemotherapy, 12 patients (11.3\%) were treated with monotherapy, 54 (50.9\%) with two drugs, 32 (30.2\%) with 3 drugs and 6 (5.7\%) with a chemotherapy scheme including four drugs. Only 2 patients received more that four different drugs. The median number of different drugs is 2.
Finally, we observed hematological recovery to ANC≥2.0×10^9/l in 88 patients (83%) with median duration of neutropenia equal to 2 days, while 25% remained neutropenic after 3 days. On the other hand, when we applied S2 to our data, we obtain an expected median duration $d_l$ of 3 days for both levels 1 and 2. Figure 2 shows the survival curves (that is, the graph depicting the number of patients remaining neutropenic as time goes by) corresponding to both groups.

**Figure 2.** Time to hematological recovery for an ANC≥2×10^9/l, for predicted low and high duration groups by score S2.

**Performance of R1 and T1 as predictors of neutropenia duration**

In Figure 3 we show the survival curves obtained for the two groups defined by R1. For patients in level 1 (<6 points) the median duration was 2 days, while for level 2 (≥ 6 points) the median duration was 3 days ($p=0.115$).

We point out that for patients where the underlying tumor was breast, the percentage with a value of R1≥6 was 71%, significantly different of the 30.7% in patients with other tumors ($p=0.000$).
Let us analyze the results for T1. In Figure 4 we present the survival curves obtained for the two groups defined by T1. For patients in level 1 (<26 points) the median duration was 2 days, while for level 2 (≥ 26 points) the median was 3 days ($p=0.002$). The percentage of patients with a breast tumor which had a value of T1 ≥26 was 25.8%, while this percentage was 61.3% in patients with other tumors (significantly different with $p=0.001$).
Because of our small sample size, we have validated our results using the resampling technique called bootstrap (see Efron [12-13], Diaconis and Efron [14], Davison and Hinkley [15]). The main idea of bootstrap is to consider the sample data as the whole population data, and to resample from this original sample in order to obtain new replicated datasets. Then, the statistical analysis will be repeated with each of those datasets. We have obtained 500 replicates, comparing in each one the survival curves for both levels (level 1=low duration, level 2=high duration). The most discriminant was the score, the highest would be the number of replicates in which the survival curves are significantly different.

Considering a significance level $\alpha=10\%$, for R1 we found that the survival curves were significantly different ($p$-value less than $\alpha$) in 39% of replicates. This percentage was only 20% for S2, while for T1 the 100% of replicates had significantly different survival curves.

**Conclusions**

Table 1 summarizes the results concerning S2, R1 and T1 for our sample. It includes the median of $d1$ (time to recovery for an ANC $\geq 2\times 10^9/l$) for the low/high level groups defined by each of the three scores, and the $p$-values for the comparisons of survival curves.

<table>
<thead>
<tr>
<th>Threshold scores</th>
<th>Low group</th>
<th>High group</th>
<th>Low group</th>
<th>High group</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Score S2</td>
<td>&lt; 8</td>
<td>$\geq 8$</td>
<td>3</td>
<td>3</td>
<td>0.458</td>
</tr>
<tr>
<td>Score R1</td>
<td>&lt; 6</td>
<td>$\geq 6$</td>
<td>2</td>
<td>3</td>
<td>0.115</td>
</tr>
<tr>
<td>Score T1</td>
<td>&lt; 26</td>
<td>$\geq 26$</td>
<td>2</td>
<td>3</td>
<td>0.002</td>
</tr>
</tbody>
</table>

**Table 1.** Median time to recovery for the different chemotherapy scores.

From the results that we have obtained, we can draw the following conclusions:

1. In our sample (from Hospital of Sagunto), the duration of CIN is usually low. For this reason, it is more difficult to discriminate between two groups. Probably, our durations are so low due to the fact that all our patients were outpatients at the onset of the FN. Then, it is possible that they arrived at the hospital when the neutropenia had already developed.

2. In our sample, score R1 discriminates better between low and high duration levels than score S2. This confirms the conclusion of Lalami et al. in [9], namely the influence that the aggressiveness of chemotherapy has over duration of FN. But on the other hand it shows that this influence is not purely additive but can be improved using different weights over the terms of the addition.
3. Our proposal T1 betters scores R1 and S2. This shows that the use of more sophisticated statistical techniques as classification trees contributes to obtain more robust predictive models. CART analysis is a powerful technique with significant potential and clinical utility. Some advantages of tree methods are: they are nonparametric and nonlinear; the final results can be summarized in a series of logical if-then conditions (tree nodes); they are very useful when there are many possible predictor variables which makes the task of variable selection difficult; they are able to model complex interactions or patterns in the data.

4. The relation between neutropenia duration and aggressiveness of chemotherapy is nonlinear and T1 has the ability to model this intrinsic nonlinearity.

We think that the results of this work highlight that the use of advanced statistical techniques for predicting neutropenia duration can lead to better results than the existing models even when using as input the same information. Although we have just a small sample, the use of the bootstrap technique makes statistically significant the validation of our results and it allows the results being extended to other potential cases from the same area. However, in order to generalize our conclusions we would need to extend our methodology to new patients from different institutions. We hope to do this in future researches.

Acknowledgements

The authors want to acknowledge the Oncology Unit of the Hospital of Sagunto for the facilities that they provided to us to carry out this work.

References


