

Model-assisted predictions on prognosis in HNSCC: do we learn?

Marc P. van der Schroeff · Kim van Schie ·
Ton P. M. Langeveld · Caspar Looman ·
Robert J. Baatenburg de Jong

Received: 2 March 2010 / Accepted: 6 April 2010 / Published online: 17 April 2010
© The Author(s) 2010. This article is published with open access at Springerlink.com

Abstract Dedicated software packages incorporating prognostic models are meant to aid physicians in making accurate predictions of prognosis. This study concerns 742 predictions of 5-year survival on consecutive newly diagnosed patients with head- and neck squamous cell carcinoma. The 5-year survival predictions made by the physicians are not compared with actual survival, but with a prediction made by OncologIQ, a dedicated software package. We used a linear regression and a linear mixed-effects model to look at absolute differences between both predictions and possible learning effects. Predictions made by the physicians were optimistic and inaccurate. Using the linear regression and linear mixed-effects models, the physicians' learning effect showed little improvement per successive prediction. We conclude that prognostic predictions in general are imprecise. When given feedback on the model's predicted survival, the accuracy increases, but only very modestly.

Keywords Head and neck oncology ·
Survival prediction · Accuracy · Learning effect ·
Dedicated software

M. P. van der Schroeff (✉) · K. van Schie ·
R. J. Baatenburg de Jong
Department of Otorhinolaryngology, Head and Neck Surgery,
Erasmus Medical Center, P.O. Box 1738,
3015 CE Rotterdam, The Netherlands
e-mail: m.vanderschroeff@erasmusmc.nl

K. van Schie · T. P. M. Langeveld
Department of Otorhinolaryngology, Head and Neck Surgery,
University Medical Center Leiden, Leiden, The Netherlands

C. Looman
Department of Public Health, Erasmus Medical Center,
Rotterdam, The Netherlands

Introduction

Prognostication remains a difficult aspect of daily medicine and oncology in particular. Due to the uncertainty of future events, physicians are often unable to give cancer patients an accurate assessment of their prognosis. This may result in non-optimal patient counselling and over- and under-treatment. In oncology the prognosis is classically based on the TNM-classification. However, it is clear that, besides TNM-classification, a variety of covariables play a role in the prognosis. The combination of factors, however, poses a difficult equation to the daily practice. Most physicians will combine their own experience and knowledge on prognostic factors in order to prognosticate. However, it is questionable how accurate these assessments are. In 2000, Christakis and Lamont [1] reported only 20% of predictions done by doctors were to be considered as accurate (predicted survival within plus or minus 33% of actual survival). This concerned 343 doctors providing survival estimates for 468 terminally ill patients admitted to five outpatient hospice programmes in Chicago during 130 consecutive days in 1996. The direction of the inaccuracy of predictions is often to the positive side, doctors are optimistic on survival prognosis [1–3]. In order to improve predictions, spreadsheets and dedicated software which present prognostic models are developed and published in literature, based on large datasets on which multivariate survival analyses are done. These programs can help physicians with patient counselling and deciding on treatment options. We hypothesise that when the clinical predictions are supported by such prognostic models, the prediction error would decrease in time.

The aim of this study is to evaluate the differences between a 5-year survival prediction done by a physician on a newly diagnosed head- and neck oncology patient

compared to the one done by a dedicated software package. We also studied a possible learning effect in the assessment of the physician if supported by this computerised prediction.

Materials and methods

This study concerns 742 predictions done by 33 physicians and the dedicated software package. The average number of predictions done by individual physicians was 22.5 (range 1–88). All 742 predictions were done on consecutive newly diagnosed patients with head- and neck squamous cell carcinoma, who were discussed in the Leiden Head and Neck Oncology Cooperative Group. The patient and tumour characteristics are given in Table 1. All participating members (otolaryngologists, radiologists, plastic surgeons, etc.) of the group were asked to make a 5-year survival prediction based on all available patient and tumour data at hand at that time. Simultaneously, these data were entered in ‘OncologIQ’, which produces a 5-year survival prediction as well. We looked at the difference between these two assessments. After each prediction the physicians were given the results of the model’s prediction in order to give them feedback. We must stress that the predictions made by the physicians are not compared with actual survival, but with a prediction made by OncologIQ.

OncologIQ is a dedicated software package, which we presented in 2001 [4].

This program is based on a Cox regression analysis on 1,396 head- and neck oncology patients. This program takes not only TNM-classification into account, but all available and relevant covariables of survival time in the Cox regression analysis. The prognostic model consists of TNM-classification, gender, age, localisation of the tumour and the absence or presence of a prior tumour.

We analysed the data in two different ways. First, we looked at absolute differences between both predictions (‘absolute residuals’). The possible decline of these differences, as a measure of a possible learning effect, was analysed using a linear regression model.

Second, we analysed the data with a linear mixed-effects model [5], in which individual physicians were declared as random factors in the model. The OncologIQ-score served as predictor, the physicians’ prediction as outcome and we used no intercept. This model then simply estimates the mean physicians’ prediction as a percentage of the OncologIQ prediction. A next step is to add the interaction of the number of successive predictions per physician and OncologIQ-score: this leads to the change in percentage over- or underestimation as a (linear) function of the successive predictions per physician. A last step sought is to differentiate these changes in time between patients

Table 1 Baseline characteristics of head- and neck oncology patient on which the predictions were made

Gender	
Male	532
Female	201
Missing	9
Age (years)	
<50	89
50–59	230
60–69	274
>70	137
Missing	12
cT-stage	
cT1	202
cT2	224
cT3	210
cT4	95
Missing	11
cN-stage	
cN0	460
cN1	63
cN2	147
cN3	53
Missing	19
cM-stage	
cM0	652
cM1	8
Missing	82
Localisation	
Nasopharynx	4
Oral cavity	125
Oropharynx	264
Hypopharynx	72
Larynx glottic	205
Larynx supraglottic	61
Missing	11
Prior malignancies	
Yes	85
No	612
Missing	45

Shown are the data entered in OncologIQ

with different characteristics as used to build up the OncologIQ-score.

For data analysis we used linear regression models in S-Plus®, version 6.

Results

Figure 1 shows the difference in predictions between the physician and OncologIQ as a function of successive predictions. The absolute difference between both predictions is on average 11%, with a range from 0 to 52%. When

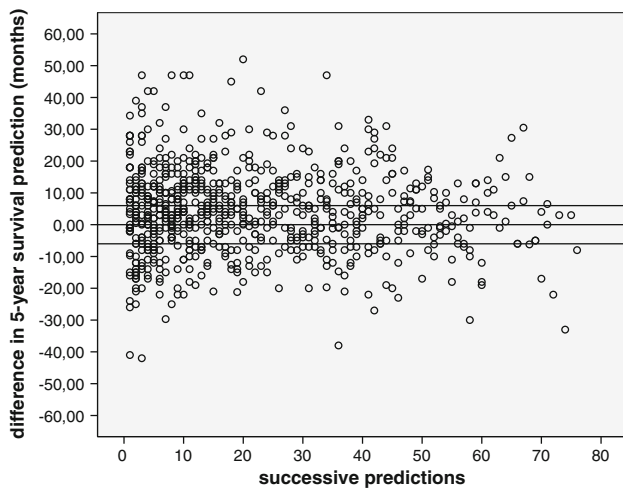


Fig. 1 The difference in predictions between the 5-year survival predictions made by the physicians and OncologIQ (reference) as a function of successive predictions. The fitted lines represent the boundaries of the ‘accurate’ prediction (6 months, 10%)

we consider a difference of $\leq 10\%$ (maximum deviation from 5-year survival prediction made by OncologIQ of 6 months) as accurate, only 277 out of 742 (37.3%) predictions classify as accurate.

Predictions made by the physicians were optimistically relative to the OncologIQ’s prediction; 459 out of 742 (61.9%) predictions made by the physicians are in absolute percentages higher than that of the program (Fig. 1).

The decline in absolute difference between the physicians prediction and that of OncologIQ was 3.6% (95% CI 0.1%, 7.1%) per successive prediction (Fig. 2). In other words, a learning effect was that the variability between physicians and OncologIQ decreases with successive predictions.

Using the linear mixed-effects model (Fig. 3), predictions from physicians were on average 4.5%, too optimistic (95% CI 2.6, 6.4%). Per successive prediction the difference between the physicians’ prediction and that of OncologIQ is declined by 0.1% (p value 0.024). A physician with more than 45 successive predictions had on average no optimism in his/her predictions compared to OncologIQ.

The last step sought is to differentiate changes in time between patients with different characteristics as used to build up the OncologIQ-score. This analysis showed no significant interactions (no data shown).

Discussion

Most previous studies on prognostication in oncology concern the prediction of survival of terminally ill cancer patients by the physician compared with actual survival

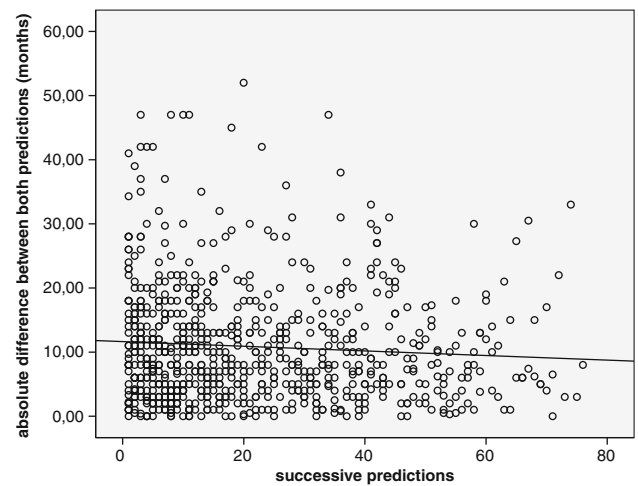


Fig. 2 The absolute difference between predictions made by physician and OncologIQ (‘absolute residuals’) as a function of successive predictions with a fitted linear regression line

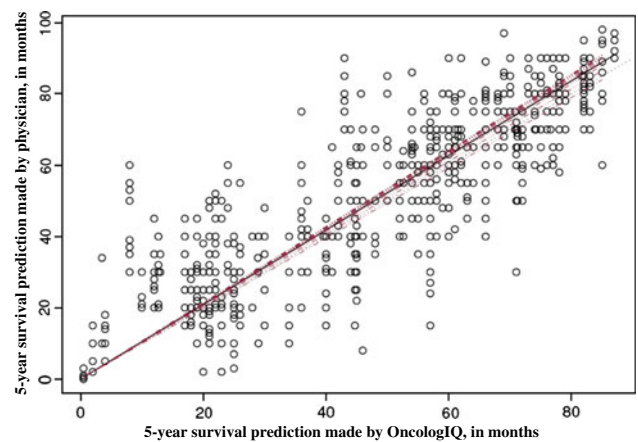


Fig. 3 The black line represents the beta of the difference in average predictions (beta 1.045). The dotted red lines represent beta’s for successive predictions. A physician with just one prediction produces a beta larger than 1.045, indicating more than average optimism. A physician with more than 45 successive predictions produces a beta around 1, indication no optimism

time and relative optimism or pessimism. Chow et al. [2] examined the accuracy of 739 survival predictions by six palliative radiation oncologists in 2004. It concerned cancer patients with metastatic disease with most common primary cancer sites being the lung, breast and prostate. The median survival of the 739 patients was 15.9 weeks. It showed that the predictions of survival tended to be too optimistic with a -12.3 weeks difference between the actual survival and the clinically predicted one. Viganò et al. [3] showed that in their study the clinical estimation of survival had a low sensitivity in terminally ill cancer patients (primary cancer sites: breast, lung, gastrointestinal and prostate) and a tendency to overestimate survival. These data concur with those of Christakis and Parkes

[1, 6] who also describe inaccurate and systemically optimistic predictions. Stockler et al. [7] studied the predicted survival in 102 newly referred patients with incurable cancer (various primary cancer sites) and found these predictions to be imprecise (29% were within 0.67–1.33 times the actual survival), but not over optimistic (35% were >1.33 times the actual survival) or pessimistic (39% were <0.67 times the actual survival). Median survival time was 12 months. Muers [8] described 196 consecutive patients diagnosed and managed as non-small cell lung cancer, who did not receive curative treatment. Physicians correctly predicted within 1 month, the survival of only 19 patients (10%). However, almost 59% (115/196) of patients had their survival predicted to within 3 months. Mackillop and Quirt [9] asked doctors to estimate the probability of cure for 98 cancer patients undergoing outpatient treatment and the duration of survival for 39 incurable patients. These patients had various primary cancer sites, including head and neck. In conclusion, the doctors were able to discriminate quite well between curable and incurable patients (area under the ROC-curve 0.91), but performed less when the duration of survival was concerned. Differences in accuracy and optimism or pessimism between these studies might be due to differences in primary cancer sites, mean length of survival and experience of the physicians. To our knowledge there has been no study published concerning the prediction of survival comparing physicians and dedicated software.

At the moment the patients in our study were discussed at the Head and Neck Oncology Cooperative Group, the data entered in OncologIQ were probably not exactly the same as the information available to the physicians. It is conceivable that some physicians (especially the one who presents the patient to the other members of the group) were aware of certain covariables that are not in the OncologIQ program. Noteworthy is the result presented by Muers [8], that a prognostic model for prediction of survival in non-small cell lung cancer patients in which the physicians' prediction of survival is incorporated showed better discriminative performance in comparison with a model without. This would also suggest that physicians are not using exactly the same factors as the prognostic model and must be using additional information. This additional knowledge, however, does not always have to be beneficial in prediction making; it could theoretically also blur the sight on more important prognostic factors.

To our best knowledge there has been no publication in which the physician's prediction is compared with that of a dedicated software program based on a multivariate survival analysis. In this way we do not compare with actual outcome (survival time), but with a maximised prediction

based on the knowledge of all relevant covariables at the time of presentation of these patients. We know from previous studies that the clinical estimation of survival is consistently imprecise. That is one of the reasons to develop these multivariable prognostic models. Our results show only 37.3% of the predictions to be accurate (maximum of 10% difference considered as accurate). We hypothesised that when the clinical prediction was supported by such a model the prediction error would decrease in time. This is, however, not that clear cut. In general, we showed little, but significant, improvement in deviation from the model's prediction with successive predictions. We therefore conclude that prognostic predictions in general are imprecise. When supported by feedback, the accuracy increases, but only very modestly. In other words we do learn, but not spectacularly. In order to maximise the patient counselling and treatment decision-making we should rely on a combination of experience and prognostic models.

Conflict of interest statement None.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

1. Christakis NA, Lamont EB (2000) Extent and determinants of error in doctors' prognoses in terminally ill patients: prospective cohort study. *BMJ* 320(7233):469–472
2. Chow E, Davis L, Panzarella T et al (2005) Accuracy of survival prediction by palliative radiation oncologists. *Int J Radiat Oncol Biol Phys* 61(3):870–873
3. Viganò A, Dorgan M, Bruera E, Suarez-Almazor ME (1999) The relative accuracy of the clinical estimation of the duration of life for patients with end of life cancer. *Cancer* 86(1):170–176
4. Baatenburg de Jong RJ, Hermans J, Molenaar J, Briaire JJ, le Cessie S (2001) Prediction of survival in patients with head and neck cancer. *Head Neck* 23(9):718–724
5. Laird NM, Ware JH (1982) Random-effects models for longitudinal data. *Biometrics* 38(4):963–974
6. Parkes CM (1972) Accuracy of predictions of survival in later stages of cancer. *Br Med J* 2(5804):29–31
7. Stockler MR, Tattersall MH, Boyer MJ, Clarke SJ, Beale PJ, Simes RJ (2006) Disarming the guarded prognosis: predicting survival in newly referred patients with incurable cancer. *Br J Cancer* 94(2):208–212
8. Muers MF, Shevlin P, Brown J (1996) Prognosis in lung cancer: physicians' opinions compared with outcome and a predictive model. *Thorax* 51(9):894–902
9. Mackillop WJ, Quirt CF (1997) Measuring the accuracy of prognostic judgments in oncology. *J Clin Epidemiol* 50(1):21–29