Salivary gland carcinoma. Stepping up the prognostic ladder
vander Poorten, V.L.M.

Citation for published version (APA):
vander Poorten, V. L. M. (2002). Salivary gland carcinoma. Stepping up the prognostic ladder
Whereas scientific knowledge is generalized and impersonal, medical practice takes place under conditions which are singular, individual and irreversible.

*Weissman, Theoretical Medicine and Bioethics, 1998*

Anyone can learn the technique, it is knowledge and judgement that makes a real surgeon.

*Michael DeBakey, 1950*
Chapter I

General introduction and outline of the thesis
General introduction and outline of the thesis

The studies presented in this thesis were performed:

1) to describe the population of patients with epithelial salivary gland carcinoma, treated in the Netherlands Cancer Institute during the period 1973-1994, with regard to treatment results,

2) to apply common epidemiological techniques for prognostic research in this population, focusing on conventional clinical and histopathological factors, thereby wanting to push the limits to a level that is as high and as practical as possible,

3) and to relate the findings under (1) and (2) to the past and current world literature on this topic.

In this introductory chapter, the specificities of the research area “malignant epithelial salivary gland carcinoma” are discussed, serving as a theoretical framework to facilitate integration of the information resulting from the prognostic studies that were performed.

In Chapter I.1, the different aspects of the domain “epithelial salivary gland carcinoma” are discussed. The picture resulting from this discussion is that of a rare disease with an extremely varied phenotypic appearance. Next, the salivary gland carcinoma patient group we examined, is shortly described.

In Chapter I.2, the epidemiological and biostatistical tools that were used to answer our basic research questions are described. Already during the discussion of this theoretical basis, some relevant links to the studies, performed and reported upon in Chapters II, III, IV, V and VI, are highlighted. We refer to these studies for specific difficulties and limitations that were encountered when applying the epidemiological and biostatistical tools to the complex study domain.

In Chapter I.3., the aims of our studies are shortly presented. An attempt to answer these questions is done by applying the methodology under Chapter I.2 to the population as described under Chapter I.1. As already pointed out in the course of Chapter I.2., the application of the methodology is mainly limited by the patient number available for study, and the number of events occurring in these patients. The analogy with “the prognostic ladder” is introduced, the prognostic researcher wanting to reach as high a step on the ladder of prognostic evidence as possible, thereby being limited by the nature of the patient group under consideration.
1. The domain “salivary gland carcinoma”

The study domain, epithelial salivary gland carcinoma, is marked by a very low incidence and a very heterogeneous appearance, both on the macro-level (anatomy) and on the micro-level (histology).

1.1. Incidence

While tumors of epithelial origin account for the majority (95%) of salivary gland malignancies, they remain an infrequent phenomenon.\(^1\) Data on annual incidence of “all salivary gland carcinomas” range from 4 to 65 new patients per 10\(^6\) people, with very high incidence rates being reported in Greenland\(^2\) and the Canadian Arctic (135 per 10\(^6\)).\(^3\) The United States incidence is reported to be 10 new patients per 10\(^6\) per year.\(^4\) For the Netherlands, different figures are available, depending on the group of patients considered.

Considering “all salivary gland carcinomas”, the incidence in the Netherlands (European standardized rate per 10\(^6\) persons) has been stable in the period 1989-1995, and was e.g. 6 for men and 5 for women in 1995.\(^5\) This is at the lower end of the rather wide incidence range as cited by Ellis and Auclair.\(^6\) Ethnicity and geographical location are thought to account for these observed differences in incidence.\(^2\) However, comparing the incidence found in the Netherlands to the incidence cited in Denmark in the period 1983-1987, being reported as 6 for men and 5 for women, it is quite likely that the Netherlands’ incidence data can be extrapolated to its neighboring countries. In 1995, 89 new salivary gland carcinomas were diagnosed in the Netherlands, at that time having a population of 15.5*10\(^6\) inhabitants (Netherlands Interdisciplinary Demographic Institute). Considered at the different echelons of health care, a general practitioner is expected to see one such patient in 50 years of practice, an otorhinolaryngologist one or two patients a year, and about ten new patients are expected to present at an average center for Head and Neck Oncology.\(^7\)

Tumors classified as “malignant epithelial salivary gland tumors” occur on different anatomical locations, and thus incidence data are also reported by anatomical site. Of the different anatomical sites, carcinomas most frequently arise in the parotid. The age-adjusted incidence of parotid carcinomas per 10\(^6\) person-years in the Southern Netherlands (Eindhoven region) in the period 1988-1992 was 3.2 for men and 1.9 for women,\(^7\) as compared to an incidence for submandibular and sublingual malignancies of 1.0 for men and 0.4 for women.
1.2. Morphology: anatomy and histology

1.2.1. Anatomy

A distinction is made between the paired major salivary glands (parotid, submandibular, and sublingual) and the minor salivary glands. The latter are the seromucous glands that are found throughout the entire upper aero digestive tract: 500 to 1000 of these glands are located in the oral cavity including lips, floor of mouth, cheek mucosa, tongue, soft and posterior hard palate, but also the nasal cavity, paranasal sinuses, nasopharynx, middle ear, Eustachian tube, oropharynx, hypopharynx and even trachea. The majority of tumors (64 to 80%) arise in the parotid glands, 15 to 32% of which are malignant. Seven to 11% arise in the submandibular glands, 41 to 45% being malignant. Less than 1% of salivary gland tumors occur in the sublingual gland, most of these (70 to 90%), however, are malignant. Minor salivary gland tumors form 9 to 23% of the entire group, one in two being malignant. This observation has been the basis for the didactic rule “the smaller the salivary gland, the less frequent a tumor arises in it, but the more frequently malignancy is involved”.

1.2.2. Histology

Histology of the normal salivary glands - histogenetical theories

The basic histology of the normal salivary gland differs according to the anatomical location and consists of different combinations of the basic components, acini on one hand and a more or less developed ductal system on the other. This can be visually appreciated in “the salivary gland unit” or “ductoacinar unit” (Figure 1).

Serous acinus cells are predominately present in the parotid and the submandibular gland, and form only a minority of acini in the sublingual gland. Mucous acinus cells are the predominate source of saliva in the latter gland, and are also well represented in the submandibular gland, and are the only source of saliva in minor salivary glands. The duct system consists of, with increasing diameter, intercalated ducts, striated ducts and excretory ducts. Luminal cells are well-differentiated epithelial cells, the abluminal part of the ducts contains (1) myoepithelial cells (also found at the abluminal side of the acini) and (2) basal cells (undifferentiated, pluripotent, according to the reserve cell theory8,9).

Two theories attempt to explain the histogenesis of this ingenious system, in an effort to shed light on both the complicated appearance of normal salivary gland tissue and also on the even more complex histological “multicellular” light microscopical phenotype of benign and malignant salivary gland tumors. The “Multicellular theory”10,11 favors a transformation of the entire
Figure 1. Histology of the normal salivary gland. One layer of myoepithelial cells invests the terminal secretory ductoacinar unit – the intercalated duct and the acinus. 


ductoacinar unit and thus requires the various "differentiated" cells to become "dedifferentiated", deranged in their growth pattern, to result in the combination of light microscopically different components observed in the various types of salivary gland tumors. The "reserve cell theory" on the other hand, states that both the normal salivary gland unit and the different tumor types are the result of differentiation of undifferentiated, pluripotent, precursor cells. Tumor cells are then supposed to arise when a problem occurs in the normal differentiation process. The earlier in differentiation the problem occurs, the more undifferentiated and "high grade" the resulting tumor type. This role of precursor cell is attributed to the "basal cells or epithelial basal ductal cells", cells that, also in the normal salivary gland, can be found at the abluminal side of the ductular system, resting on the basement membrane. In its original form, the "reserve cell theory" was called the "bicellular" reserve cell theory, distinguishing two types of reserve cells, one at the abluminal side of the intercalated ducts (intercalated duct reserve cell) and another at the abluminal side of the excretory ducts (excretory duct reserve cell). Different tumor types were supposed to result from the
intercalated duct reserve cells (adenocarcinoma, acinic cell carcinoma, adenoid cystic carcinoma, mixed malignant tumor, oncocytic tumors) on the one hand, and from the excretory duct reserve cells (mucoepidermoid carcinoma, squamous cell carcinoma) on the other. Recent research from the University of Leuven, Belgium, has furnished strong evidence for the latter theory, by indicating the common origin of both the epithelial, myoepithelial and the mesenchymal components of pleomorphic adenoma in exactly these epithelial basal ductal cells. In this way it seems possible that the scala of different histologic types of salivary gland tumors can be traced back to the common precursor cell and errors in its differentiation process towards the different normal components of the salivary gland unit.

Histological classification and grading of malignant epithelial salivary gland tumors

Non-epithelial tumors found in the salivary glands are not considered in our current work and include mainly hemangioma, lymphangioma, hemangiopericytoma, schwannoma, lipoma, granular cell tumor, fibroma and sarcoma.

a. Classification of malignant epithelial salivary gland tumors

Whatever the mechanism of tumor-histogenesis, an extremely varied picture of possible tumor types of epithelial origin is found in the clinical reality. For malignant epithelial salivary gland tumors, the 7 types distinguished in the 1972 World Health Organization Classification are displayed in table 1.

Table 1. WHO classification of epithelial salivary gland tumors

<table>
<thead>
<tr>
<th>Tumor type: carcinomas</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acinic cell carcinoma\textsuperscript{a}</td>
</tr>
<tr>
<td>Mucoepidermoid carcinoma\textsuperscript{b}</td>
</tr>
<tr>
<td>Adenoid cystic carcinoma</td>
</tr>
<tr>
<td>Adenocarcinoma</td>
</tr>
<tr>
<td>Carcinoma in pleomorphic adenoma</td>
</tr>
<tr>
<td>Squamous cell carcinoma</td>
</tr>
<tr>
<td>Undifferentiated carcinoma</td>
</tr>
</tbody>
</table>

\textsuperscript{a}In the original publication the definition is "Acinic cell tumour", and the proposition is only to use the term carcinoma, if the tumor "happens to metastasize".

\textsuperscript{b}The same holds for mucoepidermoid carcinoma, the original term published in 1972 being "Mucoepidermoid tumour". In the years following publication, with increasing clinical experience, the term "carcinoma" became the standard term.
This classification still contains the core of the most frequently encountered salivary gland malignancies and is rather limited as compared to the currently used 18-tiered 1991 WHO classification. Therefore, the 1972 classification is still appealing to a lot of clinicians, and is still widely used. Table 2, modified from Seifert and Sobin,\textsuperscript{16} displays the most recent WHO classification, marked by an increase in the number of consistently identifiable types.

**Table 2.**

The WHO 1991 histologic classification of malignant salivary gland tumors\textsuperscript{16}

<table>
<thead>
<tr>
<th>Tumor type: carcinomas</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Acinic cell carcinoma</td>
</tr>
<tr>
<td>2. Mucoepidermoid carcinoma</td>
</tr>
<tr>
<td>3. Adenoid cystic carcinoma</td>
</tr>
<tr>
<td>4. Polymorphous low-grade adenocarcinoma (terminal duct adenocarcinoma)</td>
</tr>
<tr>
<td>5. Epithelial myoepithelial carcinoma</td>
</tr>
<tr>
<td>6. Basal cell adenocarcinoma</td>
</tr>
<tr>
<td>7. Sebaceous carcinoma</td>
</tr>
<tr>
<td>8. Papillary cystadenocarcinoma</td>
</tr>
<tr>
<td>9. Mucinous adenocarcinoma</td>
</tr>
<tr>
<td>10. Oncocytic carcinoma</td>
</tr>
<tr>
<td>11. Salivary duct carcinoma</td>
</tr>
<tr>
<td>12. Adenocarcinoma</td>
</tr>
<tr>
<td>13. Malignant myoepithelioma (myoepithelial carcinoma)</td>
</tr>
<tr>
<td>14. Carcinoma in pleomorphic adenoma (malignant mixed tumor)</td>
</tr>
<tr>
<td>15. Squamous cell carcinoma</td>
</tr>
<tr>
<td>16. Small cell carcinoma</td>
</tr>
<tr>
<td>17. Undifferentiated carcinoma</td>
</tr>
<tr>
<td>18. Other carcinomas</td>
</tr>
</tbody>
</table>

This 18-tiered system has proven difficult to work with in clinical practice, and, time evolving, the picture is becoming even more complicated. First, the number of consistently identifiable light microscopically diverse types is still increasing. Among other new entities are hyalinizing clear cell adenocarcinoma,\textsuperscript{17-20} oncocytic mucoepidermoid carcinoma,\textsuperscript{21} and even primary chondrosarcoma.\textsuperscript{22} Given the recent evidence that the abluminal epithelial basal ductal cells can also differentiate into mesenchymal cells,\textsuperscript{13} explaining the “chondroid part” observed in pleomorphic adenoma, the primary chondrosarcoma is one manifestation obscuring the clear borders between epithe-
Chapter I

Salivary gland carcinomas and non-epithelial primary malignant tumors found in the salivary glands.

Second, in several histological types that classically were associated with a specific degree of clinical aggressivity, subtypes with exactly the opposite clinical behavior are described. This holds true for some allegedly low grade tumors, where subgroups with aggressive behavior have been defined. A subgroup of solid-microcystic acinic cell carcinomas shows a definitely poorer prognosis. Also the low profile aggressiveness of epithelial myoepithelial carcinoma has to be reconsidered in view of series with a recurrence rate of 40% and a disease specific survival of 60%. The "goblet cell aggressive variant" is a seemingly microscopically low grade mucoepidermoid carcinoma with a particularly aggressive behavior. Conversely, in some allegedly high-grade tumors, variants with a more indolent clinical course have been recognized. In salivary duct carcinoma (SDC), uniformly considered to be highly aggressive, a low-grade variant exhibiting the typical microscopical SDC features has been described.

To facilitate clinical use of the information contained in histological typing, the different tumor types are frequently reclassified according to their clinical behavior. Several systems have been proposed, most of them being two-tiered or three-tiered. Aggressive tumors are then called "clinically high grade tumors", an "intermediate grade" group is sometimes considered, and the slow growing, indolent, rarely metastasizing tumors are described as "clinically low grade tumors". It is important to distinguish this "grade" corresponding to clinical behavior, which creates a subdivision between histological types, from histological grading, corresponding to morphological features associated with degree of differentiation and invasiveness of growth, creating a subdivision within histological types. The fact that clinical behavior and "clinical grade" often parallels histological grade, explains the frequent mix-up. In quite some publications, the inexplicit use of these two types of grading creates confusion. Table 3 shows the still widely used dichotomy in the 1972 WHO classification.

<table>
<thead>
<tr>
<th>Low-grade</th>
<th>High-grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acinic cell carcinoma</td>
<td>Adenoid cystic carcinoma</td>
</tr>
<tr>
<td>Low-grade mucoepidermoid carcinoma</td>
<td>Squamous cell carcinoma</td>
</tr>
<tr>
<td></td>
<td>Adenocarcinoma</td>
</tr>
<tr>
<td></td>
<td>Carcinoma ex-pleomorphic carcinoma</td>
</tr>
<tr>
<td></td>
<td>High-grade mucoepidermoid carcinoma</td>
</tr>
</tbody>
</table>

Table 3. Dichotomy into low-grade and high-grade salivary gland carcinoma

20
Table 4 shows a more recent proposal for a three-tiered subclassification in the 1991 classification as proposed by Klijanienko and Vielh (modified) from the Institut Curie in Paris.\textsuperscript{35}

\textbf{Table 4.}

\begin{center}
\textbf{Three-tiered subclassification in the 1991 classification as proposed by Klijanienko and Vielh (modified)}\textsuperscript{35}
\end{center}

\begin{tabular}{|l|l|l|}
\hline
Low-grade & Intermediate-grade & High-grade \\
\hline
Polymorphous low-grade adenocarcinoma & Mucoepidermoid carcinoma & Adenoid cystic carcinoma \\
Metastasizing pleomorphic adenoma & Adenocarcinoma NOS & Carcinosarcoma \\
Basal cell adenocarcinoma & Carcinoma ex pleomorphic adenoma & Squamous cell carcinoma \\
Acinic cell carcinoma & Lymphoma & Undifferentiated carcinoma \\
& & Salivary duct carcinoma \\
& & Myoepithelial carcinoma \\
& & Epithelial-myoepithelial carcinoma \\
& & (Papillary) cystadenocarcinoma \\
& & Sebaceous carcinoma \\
& & Mucinous carcinoma \\
& & Oncocytic carcinoma \\
& & Secondary tumours \\
\hline
\end{tabular}

It is this type of “reduction according to clinical behavior” that is commonly used in prognostic factor studies, and this is also the form in which we have analyzed the variable “histological type”.\textsuperscript{36-38} For a rare disease like salivary gland carcinoma, this “level reduction” is obviously interesting from the numerical point of view. For statistical computation, it is clearly more interesting to have to subdivide a small patient sample in only two or three categories than to have to consider 18 categories. It has to be realized, however, that a lot of interesting information contained within the more detailed classification is not considered in this way.

b. Grading of malignant epithelial salivary gland tumors

This aspect of the histopathological work-up has already been touched in the previous paragraph. Histopathological grading attempts to explain variable biologic behavior within one histotype, but remains controversial because of lack of uniformity and insufficient independent prognostic power. Uniformity remains problematical\textsuperscript{39,40} due to a weak consensus on grading criteria. Multiple collinear-
ities (stage-grade\textsuperscript{41}/age-grade\textsuperscript{38}/irradical resection-grade\textsuperscript{42}) often make grading a non-essential piece of information in clinical decision making. Generally, a grading effort is done in three histopathological types: mucoepidermoid carcinoma, acinic cell carcinoma and adenoid cystic carcinoma.

In grading mucoepidermoid carcinoma, two and three-tiered systems have been advocated. Batsakis, et al.\textsuperscript{40,43} follow the suggestion of Healey, et al.\textsuperscript{44} for a three-leveled system, based mainly on the involved cell types. This system is supported by univariable survival and cure effects in Healey’s series, where recurrence rates of both low and intermediate grade carcinomas are quite close together. Healey, et al.\textsuperscript{44} describe a recurrence rate of 6\% and 20\% for low and intermediate grade mucoepidermoid carcinomas respectively, vs. 78\% for high grade, and disease specific death rates in 6\% and 8\% for low and intermediate grade mucoepidermoid carcinomas, respectively, vs. 72\% for high grade lesions. Batsakis and Luna\textsuperscript{43} also cite Spiro, et al., who used comparable grading criteria,\textsuperscript{45} but they wrongly cite a 83\% cure rate for the intermediate group in the series of Spiro et al, whereas the authors mention a 63\% cure rate.\textsuperscript{43} The Armed Forces Institute of Pathology (AFIP)\textsuperscript{41} also favors a three-leveled grading system, based on different criteria, and promotes a more quantitative approach. We, however, do not feel, that the clinical data supporting this complicated grading system, are very convincing.\textsuperscript{41} Brandwein, et al.\textsuperscript{27} showed a good correlation of these AFIP grades with stage and subsequent local control, but revealed a significant grading disparity, even using the “standard AFIP grading criteria” among 5 participating experienced oral pathologists. They formulated yet another grading proposal. More appealing to clinicians are simpler two leveled grading systems,\textsuperscript{16,46} using less and less complicated criteria (Evans: \(>90\%\) solid tumor = high grade – only univariable analysis to support this ; Seifert: \(>90\%\) epithelial cells = high grade). Spiro\textsuperscript{45} in his original study also joins intermediate and high grade in his analysis, because of the markedly more aggressive behavior and the need for maximum (combined) treatment in these two subgroups. Only the low grade subgroup is manageable with surgery alone.

For acinic cell carcinoma, the empiric three leveled grading system as proposed by Batsakis, et al.,\textsuperscript{47} combines histopathological and clinical features. Until now, no hard data support the usefulness of this grading system. Here also, the collinearity with disease stage seems to overrule the information resulting from histopathological grading.\textsuperscript{48} The inclusion of tumor size in the histological grading criteria exemplifies and induces this collinearity. Interesting, however, is the focus this effort brings on a high grade subgroup of acinic cell carcinoma where even in low stage combined treatment should be considered. This point is also made in a recent large overview\textsuperscript{49} of patients from the National Cancer Data Base. A subset of acinic cell carcinomas
(16.5% of 438 graded tumors, out of a total of 1353 patients with acinic cell carcinoma) has a propensity for being larger and having neck metastasis at presentation, and (often irradical) resection should be followed by radiotherapy.

Every adenoid cystic carcinoma (AdCC) should by all means be considered a "clinically high grade malignant neoplasm". The three-leveled grading system\(^50\) into tubular (grade I), cribriform (grade II) and solid (grade III) subtypes univariately explains differences in tumor control,\(^51,52\) but will not change the clinical management. The three patterns can be present in one tumor, and especially tubular and cribriform growth patterns frequently overlap.\(^53\) Spiro and Huvos found no definite prognostic role for histopathological grading of this tumor type, prognosis being mainly determined by tumor stage.\(^53\) A recent interesting study confirmed these findings regarding stage, and found an additional independent role for immunohistochemical detection of e-cadherin, reduced expression paralleling increasing grade. This approach, when confirmed, may be a more direct measurement of aggressivity than optical grading.\(^54\)

Whereas occurrence of distant metastasis in AdCC is currently reported to occur in around 40% of the patients and to be related to initial stage of the tumor\(^53\), an exceptionally high rate was found in an interesting study of AdCC showing a cumulative incidence of distant metastasis in 100% of patients followed up for longer than 5 years, regardless of grade and initial stage.\(^55\) Tumor doubling time studies by the latter authors suggest that distant metastasis would very often be present subclinically long before the initial diagnosis of the primary salivary gland tumor.

The only grading that has clinical treatment influencing implications until now is the grading of mucoepidermoid carcinoma into low grade versus intermediate and high grade types, the former requiring no additional radiotherapy in completely resected Stage I and II tumors.\(^56,57\)

1.3. The study population

1.3.1. The Netherlands Cancer Institute: salivary gland carcinoma patients 1973-1994

In the period 1973-1994, 326 patients with the diagnosis "salivary gland carcinoma" were treated at the Netherlands Cancer Institute. Of these, 53 patients presented with a recurrent or metastatic tumor following treatment of their primary tumor in another hospital, whereas 273 patients were diagnosed with and subsequently treated for a primary salivary gland carcinoma.
According to subsite, 43 patients (15.8%) were diagnosed with a primary submandibular gland carcinoma, 55 patients (20.1%) were diagnosed with a carcinoma arising in the minor salivary glands, 168 patients (61.5%) presented with a primary parotid carcinoma. Only 7 patients (2.6%) presented with a carcinoma of the sublingual gland. The clinical and histological characteristics of the patients with primary submandibular carcinoma are extensively described in the “patients” section of Chapter II, and the same holds for the patients with minor salivary gland carcinoma (Chapter III) and parotid carcinoma (Chapter IV).

1.3.2. The Dutch Head and Neck Oncology Cooperative Group parotid carcinoma patients 1985-1994

The Dutch Head and Neck Oncology Cooperative Group (NWHHT) is an organization with the aim of improving research in Head and Neck Cancer by joining efforts among tertiary referral centers. One of its projects was and still is creating and updating databases of patients with Head and Neck Cancer treated in its member centers. As a member of the Dutch Cooperative Group, we had the opportunity to use the database of patients treated for primary parotid carcinoma in the Netherlands Cancer Institute (NKI/AvL) and in the academic centers of Groningen (AZG), Leiden (UMC Leiden), Maastricht/Heerlen (AZM/RTIL), Nijmegen (AZN), Rotterdam (AZR/DDHK) and Utrecht (UMC Utrecht) from 1985 to 1994. The original database consisted of 333 records from consecutive patients treated for parotid carcinoma. Of these 333, we excluded 7 records. Four patients with unknown tumor status at the end of follow-up after 1, 6, 11 and 88 months, for whom date of last tumor status could not be reconstructed, were excluded. Two patients both had 2 records. For the first patient, the 2 records were almost exactly the same, except that follow-up was more complete in one record. The less complete record was excluded. The two records for the second patient had a number of differences, indicating a double primary tumor. Both records were excluded. The total number of patients valid for consideration was thus 325. For our major computations, and as explained in Chapters V and VI, this group was further reduced to 231 patients, after leaving out the patients treated in the Netherlands Cancer Institute (NKI/AvL), and the patients not free of disease at the end of treatment. The latter patient group is extensively described in Chapters V and VI.
2. Prognostic research. Epidemiological and biostatistical aspects, application to our domain

2.1. Definition of prognosis

"The essentials of clinical practice of medicine include diagnosis, therapy, prophylaxis, and prognosis. The first three are based on action. The last essential, prognosis, is an art and a science of prediction and is based on the knowledge distilled from the diagnosis and the information gained from prior experience".58

Prognosis refers to the possible outcomes of a disease and the frequency with which these can be expected to occur.59 Relating the heterogeneity in the outcome of disease to (one or more) different variables measured in individual patients is the object of the study of prognostic factors.60 Reasons to define this relationship between prognostic factors and outcome, are, as stated by George60 and Altman and Royston,61

a. to inform treatment or other clinical decisions for individual patients. In the field of salivary gland cancer this could e.g. be to avoid exposing low-risk patients (as defined by prognostic factors present at the moment of primary surgical treatment) to the strains of an adjuvant radiotherapy or an elective neck dissection;

b. to inform patients and their families regarding the potential outcome of the disease. This is a medico legal responsibility, more applicable in systems where informed consent is established (US), but also in Europe an increasing demand for quantitative information on outcome for a given disease treated by a given treatment is observed;

c. to create clinical risk groups for stratifying patients by disease severity (in clinical trials). This is essential for observational studies but may also apply to clinical trials, in order to improve comparison of patient groups for treatment effect, either by stratified randomization or adjustment during analysis. However, to date, no such trials have been performed in the field of salivary gland cancer. Prognostic information may also help to select patients, mostly the ones with a bad prognosis under the current treatment types, to plan future studies with new treatment types: creating risk groups helps in defining eligibility criteria.62

d. understanding of the disease process, learning about the "treated history" of a disease – provides directions for further study.

A continuing interest in prognostic factors is exemplified by a number of publications addressing this cornerstone of medicine.58,63-66 Several
researchers from different fields in medicine consider the development and validation of prognostic indices the way to go for an improved and acceptable prognostication.\textsuperscript{63,67-70}

2.2. Study of prognosis in oncology

Prognostic research in oncology typically tries to predict specific outcomes (see 2.2.1) by relating them to specific prognostic factors (2.2.2), using specific techniques (2.2.3). This chapter deals with these basic notions. In the margin of this discussion, links to the studies adopted in this thesis are made, discussing the limitations when using the theoretical base in the practice of prognostic research for salivary gland carcinoma.

2.2.1. Outcomes in oncology: which one to study, what are the specificities?

The most frequently studied outcomes in oncology are usually time-event outcomes, consisting of a time variable (i.e. an interval between two dates) and an indicator variable, indicating whether the event under study (death, recurrence, complications, etc.) occurred at that time or not. This event is commonly called "failure" for it is exactly the opposite of what the oncologist and the patient are trying to achieve. The time factor indicates the time at which the event occurs or the time at last follow-up. In choosing between the different events, there is always a trade-off between relevance and reliability. For four events, a definition will be followed by remarks on relevance and reliability.

a. Overall survival: very reliable – less relevant

When considering this outcome, every patient that has died at the moment of study closure, from whatever cause, is considered a treatment failure. This results in a decline in the survival curve at every moment in follow-up a patient dies.

A lot of non-tumor related factors (intercurrent disease, psychological status, age, gender, etc.) can influence this outcome, and make this outcome less interesting for evaluation of treatment results and the prognostic factors associated with these results.

An advantage of this outcome and the reason why studies do use it frequently, is that it is more reliable and easier to determine than e.g. tumor recurrence. The investigator can simply use the death certificate, or call the general practitioner in case of losses to follow-up, and still get a reliable statement on the survival status and the date of death.
b. Disease specific survival (DSS): less reliable – very relevant

Only patients dying as a direct result of tumor are considered treatment failures in this instance. This is a very interesting measure, as intercurrent diseases leading to patients’ death do not influence the height of the survival curve, and hence the observed effects of the treatment and the prognostic variables under study on tumor control. The disease specific survival curve, by definition, should be located at the same (if patients die only because of their tumor) or mostly at a higher level than the corresponding overall survival curve. The interesting point when using this outcome, is that patients who suffer tumor recurrence, but are cured subsequently (resection of local or neck recurrence, resection of isolated lung metastasis), are not considered treatment failures. It is thus a measure of “best obtainable treatment result” of “all possible treatment given”, including salvage treatment for recurrence. The disadvantage is that the cause of death is always much more difficult to assess than death itself; here lies a problem of “assessment bias”. Furthermore, of deceased patients that are judged to be disease free, the investigator never knows to what extent disease was latently present, or what the chances of recurrent disease would have been, had the patient not died. However, when using methods like Kaplan-Meier analysis (see 2.2.3), the latter is not a problem if the death is unrelated to the existence of the tumor.

c. Disease free survival (DFS)

For this outcome, to be considered as treatment failure, patients must have had a tumor recurrence or have died at the moment of evaluation. This measure suffers the same flaw as the outcome “overall survival”, in that a lot of non-tumor related factors can influence this outcome. It is, however, a frequently used outcome in oncologic prognostic research, but many authors confound it with the next outcome to be discussed, disease free interval. This mistake can be seen in any study presenting a DFS curve that is located at a higher level than the overall survival curve, whereas this logically should be lower, as also patients that are alive after a tumor recurrence are considered treatment failures. This is a realistic situation in patients with salivary gland carcinoma: a lot of patients are alive for years with distant metastasis (e.g. typical situation in adenoid cystic carcinoma). They thus lower the observed disease free survival curve, whereas they are considered “treatment successes” when looking at overall survival. The latter curve should thus, by definition, be located on a higher level than the corresponding DFS curve.

d. Disease/recurrence free interval (DFI or RFI)

General RFI looks at all possible sites of recurrence and is mainly interesting
from the point of view of the newly diagnosed patient, answering the question: how is the chance of being cured following this first therapy? Provided there is an adequate follow-up and a substantial effort to detect tumor recurrence, this measure can be considered the most direct reflection of treatment success in terms of "cure following the initial treatment". A subdivision according to the site of recurrence (local, regional, distant) can be made, mainly interesting to evaluate the effect of locoregional treatment. This can be interesting in case of salivary gland carcinoma, as the two treatment types used, surgery and radiotherapy, are locoregional treatment modalities.

At this moment, however, the problem of competing risks arises. It is likely that the chance of occurrence of distant metastasis is not independent of the risk of developing locoregional recurrence, and an early death may render assessment of treatment effect on locoregional control impossible. A further discussion of this more difficult topic is beyond the scope of this thesis.

For the latter two outcomes (DFS and RFI) to be reliable, all patients should be examined consistently at fixed times after entering the study. This prerequisite was no problem in the patient group of the Netherlands Cancer Institute, who adhere closely to a fixed follow-up protocol: patients are reevaluated every 4 weeks during the first year following treatment, every 2 months during the second year, every 3 months during the third year, every 4 months during the fourth year, every 6 months during the fifth year, and yearly from the sixth year on.

For all four measures starting and end date of the interval also have to be defined clearly. The starting date should be chosen as the date at which inclusion criteria for the study are first satisfied. If we want to include only surgical patients, then the time of surgery should be the starting date. However, if we want to include all patients with a certain diagnosis, then date of diagnosis is the relevant starting date. If an event during follow-up (e.g. metastasis) defines inclusion, then that date should be the starting date. Regarding the end date, it is always wise to choose a fixed closure date for the study and try to follow-up all patients up until that closure date. Later events should be disregarded in order to prevent overestimation of the number of events. Events are often more likely to come to the attention of the researcher than uneventful follow-up visits.

We studied overall survival, DSS and recurrence free interval and the prognostic factors determining these outcomes for patients with submandibular carcinoma (Chapter II) and minor salivary gland carcinoma (Chapter III). The multivariable analysis for patients with parotid carcinoma (Chapter IV) considered only RFI, for it was our aim to devise a patient-informing model,
and it was our belief that the question "What are my chances of being tumor-free following this first treatment" was the most relevant one in this respect.

2.2.2. Prognostic factors

For oncological patients in general, but specifically for patients with salivary gland carcinoma, a remarkable heterogeneity in the outcomes described under 2.2.1, can be observed. A logical next step is to go and search for "determinants" of these diverging outcomes. The term "determinant" should in this respect not be considered a causative factor, but merely a strongly associated factor. Such determinants are commonly called "prognostic factors", variables that can account for some of the heterogeneity associated with the expected course and outcome for a patient with a specific disease, and are commonly looked for in characteristics of the particular patient or of the disease he is affected by. In this respect prognostic factors are usually distinguished from "risk-factors", patient characteristics associated with the development of the disease in the first place, implying a more causative relationship.

Prognostic factors can be of several types:

a. Semantic classification

In general prognostic research, demographic (age, gender, race, etc.), disease-specific (cardiac status, tumor stage, etc.) or comorbid (performance status, etc.) prognostic factors are distinguished. Oncologic prognostic research for rare diseases such as salivary gland malignancies, often focuses only on demographic and disease specific prognostic factors. Distinguished categories are patient-characteristics (age, gender, symptoms, signs), tumor-characteristics (clinical, histopathological, and lately molecular biological characteristics), and treatment characteristics.

b. Classification according to behavior in time

Distinction is made between time-dependent and "fixed" covariates or prognostic factors. Fixed prognostic factors such as gender, age at diagnosis, etc., do not change during follow-up, while time-dependent factors such as the value of serial biological markers or response status do. In our studies, all factors examined are fixed. Specialized statistical methods have to be used to deal with time-dependent factors.

c. Classification according to definition of scale of a prognostic factor

Prognostic factors can be categorical (ordered, e.g. T and N classification vs. unordered, e.g. histological types) or continuous (e.g. age at diagnosis).
Continuous variables can be used as such in multivariable Cox proportional hazards analysis (Chapter IV and V), but have to be reduced to ordered categorical variables for univariable analysis using the product-limit method of Kaplan and Meier (Chapter II and III). Both ordered categorical and continuous variables should be handled with proper care in statistical analyses.

In practice, prior knowledge is used to indicate which are relevant factors to consider. This usually results in a fairly limited number of candidate variables. Our studies were all set up to confirm or reject candidate prognostic factors that were identified earlier in the literature. A table displaying the variables studied and the scale upon which they were defined is given in each chapter.

2.2.3. Techniques to study the relationship prognostic factors – outcome

a. Clinical epidemiological aspects

a.1. Study design

The best study to identify and quantify the effect of one or more prognostic factors is a prospective cohort study. In such a study, a well-defined group of patients (in casu newly diagnosed patients with a primary salivary gland carcinoma), not yet having suffered the adverse outcome, has to be followed over time for an adequate length of time. At the beginning of follow-up, patients differ regarding baseline values of essential known prognostic factors and also hypothetical prognostic factors. These baseline values have to be registered, and the prospective design is ideal for minimizing missing values in this essential information. Follow-up time elapsing, failures are registered, preferably using a rigid follow-up scheme and standardized detection methods.

The relatively rare incidence of salivary gland malignancies and the well-known long term evolution, with some patients recurring only after 15 years, has until now precluded this type of study. For prognostic research, we are forced to rely on “historical cohort studies” or retrospective studies, which are much more vulnerable to bias and missing data, because they depend on the accuracy of the medical records.

The studies we performed were for this very reason essentially all “historical cohort” or “retrospective studies”. Several techniques to optimize follow-up (written contact with general practitioners for data regarding tumor status or death) and to minimize effect of missing data (use of dummy variables in multivariable analysis) were used to overcome these problems as adequately as possible.
a.2. The prognostic ladder: increasing evidence parallels increasing complexity of studies

A helpful classification of studies of prognostic factors is the one proposed by Simon and Altman, a classification which parallels clinical trial terminology. Mainly considering the advent of new, molecular biological factors that allegedly carry prognostic information, Simon and Altman distinguish between:

1. Early exploratory studies (phase I) are just checking whether a specific factor is associated to the outcome of interest. The studies described in Chapter II and Chapter III can be considered early exploratory studies, in that they only verify the described association of certain reported prognostic factors, without real quantification of effect.

2. Extended exploratory studies (phase II) generate hypotheses from extensive analysis of data. The study described in Chapter IV can be considered a phase II, extended exploratory study, screening for interesting hypotheses to be tested in independent data sets. The prognostic index resulting from it can be considered the scientific hypothesis, stating a testable quantification of the different identified prognostic factors.

3. Large confirmatory studies of prestated hypotheses (phase III studies). Most of these are studies linking prognosis and therapy, defining subsets of patients benefitting from a given therapy, allowing for precise quantification of the observed effect. Notwithstanding the absence of a link to therapy, the study in Chapter V is a real confirmatory study of a hypothesis stated in advance (the prognostic index generated in Chapter IV) and can thus be considered a phase III prognostic factor study.

a.3. Quality checkpoints to judge a prognostic factor study

Several of these checkpoints can be found in Table 1 in the article by Simon and Altman “Guidelines for phase III prognostic factor studies”. However, this table mainly aims at judging prospective prognostic factor studies examining the additional effect of a potentially new, often molecular biological factor, to the effect of a set of “standard prognostic factors”. For this reasons not all of the points mentioned in that table can be extrapolated to judge our studies. Other "quality” points that are discussed are drawn from Laupacis, et al. and George.
a.3.1. Clear inception cohort

An important prerequisite is that “a clear inception cohort should be assembled”, with few patients non-evaluable due to missing data (<15%, <20%).

A representative and well-defined sample of patients at a similar point in the disease follows clearly defined inclusion and exclusion criteria, so that later on it is clear to which patients the results can be extrapolated.

The patients we are working with in our studies (Chapter II, III, IV, V, VI) are all patients with their primary treatment in a tertiary referral center, so that conclusions will be applicable to patients in this setting. Throughout the world, most patients with the rare condition “salivary gland carcinoma” are predominately referred to tertiary centers for treatment.

Values can be missing with regard to outcome and with regard to modifying variables.

Missing values regarding “outcome” can be minimized using the strategies as described under a.1. Optimizing conclusions from a patient series with remaining missing values requires use of the survival analysis technique, which takes into account “censored observations” (see 2.2.1) in reaching treatment outcome estimates.

The problem of missing values regarding “modifiers of this outcome”, i.e. prognostic factors, can affect univariable analysis badly and multivariable analysis even worse.

The simplest way to tackle this problem is to eliminate all subjects with missing values from the analysis. This is only valid if the missing data are missing completely at random, meaning that patients with missing data do not differ systematically from the other patients regarding the prognostic factors-outcome relationship examined; and this may be a difficult assumption to support. Deletion of patients with missing values of course also decreases precision and power of the study. In the studies in Chapter II and III, patients with missing data were simply deleted from the analysis. The assumption of “data missing completely at random” seems hard to hold if too much values are missing. A rule of thumb proposed by George is that covariates with at least 20% of missing values are to be eliminated and not considered for analysis. In this way, e.g. in the multivariable analysis in Chapter IV, T status, associated with the largest number of missing values (missing in 17% of patients), is retained in the analysis (Table 3 of Chapter IV).

If this “deletion of patients with missing values” is to be done in multivariable analysis, the effect is multiplicative, and soon no data remain for analysis. George describes a hypothetical situation where in a study with 18 factors to evaluate, and every factor randomly missing in 10% of patients, only 15% of patients remain with information on every covariate. A mathematical
rescue mechanism, which we have applied in Chapter IV, is the use of “dummies” in multivariable analysis. In this strategy, dummies (0 = unknown/not applicable, 1 = known) were added to the model when evaluating factors with missing values and factors referring only to subcategories of the total cohort. The factors themselves were assigned some arbitrary value, preferably the reference level of the observed values. As an example of the application of the “dummy” principle, the estimate of the prognostic impact of the factor “presence of perineural growth” can be considered. This aspect can only be determined in the subcategory of surgical patients, where a resected specimen can be examined for this feature. It makes only sense to estimate the prognostic power of this factor in the subcategory of patients where this information is not missing: those who have undergone surgery and where presence of perineural growth was determined. Patients with code 0 for the dummy are assigned to the reference category of the variable involved, and both the factor and its dummy are entered simultaneously in the multivariable analysis. In this way patients with missing values do not contribute to the estimation of the prognostic power of the missing variables themselves, but are still used for estimation of the prognostic power of other variables. For this method to work, it is still required that missing information is completely at random. More advanced methods of dealing with missing values such as multiple imputation or maximum likelihood estimation using a joint distribution for the prognostic factors and the response are still difficult to use in standard statistical packages, such as SPSS and SAS, available to us at the time of analysis. They are also no panacea for missing values as they generally have to rely on untestable assumptions.

a.3.2. The hypothesis should be stated in advance

The hypotheses of the studies in Chapters II, III and IV are the list of variables that were selected from the available information in the literature, and that are displayed in a table in every study. The hypothesis at start for the studies in Chapter II and III was “the variables listed in the table have a uni-variable prognostic power for the different outcomes considered”. Only in the study on parotid carcinoma, the number of patients studied was sufficient to allow for a subsequent multivariable assessment of prognostic ability. The resulting prognostic index served as the basis for the new hypothesis to be tested in the study in Chapter V. The hypothesis here was that a good prognostic ability would be retained in independent patient material.

a.3.3. Follow-up should be sufficiently long and complete

The reason a patient is lost to follow-up may well be related to outcome. Patients may fail to return because they have suffered exactly those events in
which the investigators are interested (they have died, a recurrence appeared for which treatment was sought elsewhere, or patients who feel cured may not want to come back for follow-up). In this respect, a proposed measure of quality for the study considered, is the number of events, compared to the number of patients not having follow-up up to five years. Taking Chapter IV as an example, 2 out of 151 patients did not have follow-up at 5 years, as compared to 57 events or recurrences, all but two occurring within 5 years of follow-up.

a.3.4. An objective outcome is needed

The respective outcomes studied for the different studies in the different chapters were extensively discussed in paragraph 2.2.1.

a.3.5. Adjustment for other prognostic factors, treatment factors

This is only applicable to the studies in Chapter IV and V, where the technique of multivariable analysis allows for adjustment of estimated impact of one prognostic factor for the effect of other prognostic and treatment factors. The number of available patients treated with submandibular and minor salivary gland carcinoma limits the researchers possibilities in the studies described in Chapter II and III.

a.3.6. Report on all results

All results, positive and negative, significant and not significant, should theoretically be reported, and in our studies they all have been reported (Chapter II, III, IV, V and VI).

a.3.7. Practicality of results

The final results should be formulated as practical as possible. This is very difficult when reporting on results of studies following only univariable analysis. Ideally a prognostic factor study should answer the individualized question “how large is the likelihood of the outcome in a specified period of time for this patient with this combination of prognostic factors”. At that time a table or formula combining the known prognostic factors in a weighted manner, allowing to individualize the statement on prognosis, comes in handy. This was the objective of the creation of a prognostic index in Chapter IV, and the subsequent transformation of this index in a computerized easy-to-use program, as added to Chapter V. The observation of the recurrence free interval curves resulting from this analysis can demonstrate the time course of the risk of the outcome to occur (constant in time, or e.g. only high initially, then remaining constant). Estimates of likelihood should be accompanied by a measure of precision of estimation (confidence intervals).
In conclusion, the reader of a study of prognostic factors needs an answer to the question: “Will the results help me in caring for my patients?”. He needs to be able to determine whether the patients that generated the hypothesis of prognostic importance of the described factors are similar to his own patient. The clear inclusion and exclusion criteria as well as the referral pattern as discussed under a.3.1 are important in this respect, as well as the presentation of the study results in the form of a prognostic formula, allowing to individualize the statement on prognosis. The information delivered in this thesis essentially cannot be expected to guide decisions about therapy. As a rule, in the management of salivary gland carcinoma, the primary treatment consists of primary surgery, in most instances followed by radiotherapy. Only rarely the latter treatment can be omitted, and although the prognostic factors described in Chapters II, III and IV can help in this decision, not one of the studies is able to deliver clear treatment guiding information because of the retrospective nature. The main use of the presented information on prognostic factors is for information of the patient and relatives. A quantified good prognosis is a good basis for reassurance, a quantified bad prognosis can help in counseling. Important in this respect is the next topic, a.3.8, the availability of evidence of generalizability.

a.3.8. External validation is needed

Once prognostic factors are identified and quantified, external validation is needed. As stated by George, “It is almost certain that the apparent strength of the relationships, and often their actual existence, will be overstated if the results are taken at face value. The proper attitude here is one of skepticism and caution, always demanding further confirmatory studies before putting much weight on the results”. Using the same data set to derive a model and to test its performance results in an inflated estimate of the effect of the retained prognostic factors. “A fair evaluation of discriminative power of a predictive model requires independent data”. This was the object of the study described in Chapter V: to establish generalizability, meaning, to prove, that the index works well for other patients. It is checking whether a good discrimination is achieved in another group of patients with the same disease, in a broader geographical and historical context.

b. Biostatistical aspects: survival analysis

The results and conclusions presented in this prognostic research oriented thesis are all reached using a specific data analysis approach called survival analysis. Survival analysis has three basic goals and is therefore the ideal tool for our purpose.
b.1. Estimating survivor functions

To estimate and interpret survivor functions from survival data. This is what we are interested in, in this salivary gland carcinoma research: treatment results in terms of survival, disease specific survival, and follow-up time without recurrence.

b.2. Univariable analysis

To compare survivor functions. We want to identify and quantify prognostic factors. Investigating prognostic factors means dividing the patient group under study in subgroups according to the levels of the prognostic factor under study and subsequently compare their respective survivor and/or hazard functions.

b.3. Multivariable analysis

To assess the coexistent relationship of different explanatory variables, in our case prognostic factors, to survivor time. This is possible in survival analysis using mathematical modeling (e.g. Cox proportional hazards model\textsuperscript{81,82}). We need this because we want to study different prognostic factors, each with their respective levels, at one time, so that we can figure out which factors are the more important ones.

At b.1. Estimating survivor functions

As extensively discussed under Chapter II.1., the outcome variable studied is “time until an event occurs”, this time reflecting the prognosis of the salivary gland cancer patient. The event that is registered during follow-up can be death from any cause (overall survival), death due to tumor (disease specific survival), or tumor recurrence (recurrence free interval). For the sake of simplicity, in the following discussion the terms survival and death are used but can be substituted by every other outcome mentioned.

In this type of analysis, the ideal situation is to have, for every patient studied, the time of death. Often, in real life, many patients are not followed long enough to have reached this endpoint and thus we know only a minimal survival time. This is e.g. the case when a patient is followed-up for some time, was disease free at last follow-up, but was subsequently lost to follow-up. Such an event is called a “censored event”. The information carried by a patient surviving before the study end is adequately used by survival analysis. This is because, in this analysis, the expected survival of a censored patient beyond the censoring time is extrapolated based on the survival experience of the patients still at risk at the censoring time. Of course, in calcu-
lating standard errors and $P$-values, the censoring is taken into account.

Survival probability at any given time in follow-up can be displayed as $S(t)$ or $h(t)$.

$S(t)$ is the survival probability at time $t$ and thus has values between 0 and 1 or between 0 and 100 when expressed as a percentage. The mathematical notation is $S(t_{(i)}) = S(t_{(i-1)}) \times \text{Pr}(T>t_{(i)} | T>_{(i-1)})$. The evolution of the survival probability in time can be graphically displayed as a survivor curve using the Kaplan-Meier product-limit method.\textsuperscript{83,84} In this way allowing for a simple visual assessment of the complete survival function of the study cohort. This way to display the survival experience is sufficient for univariable analysis.

Another way to describe the same survival experience of a patient cohort is the hazard rate, $h(t)$. $h(t)$ is also called a conditional failure rate. The mathematical form $h(t)=S(t)^{-1}/S(t)$, where $S(t)^{-1}$ is the slope of the survival curve at time $t$, can be interpreted as an instantaneous probability per unit time for the undesired event to occur, given this event has not occurred until $t$. $h(t)$ is a rate and thus has values ranging from 0 to infinity, this value depending on the time unit used. This form of survival experience is the basis of the Cox proportional hazards analysis.

At b.2. Univariable analysis: to compare survival according to two or more levels of one variable\textsuperscript{84}

The first step in evaluating a prognostic factor is to see whether different outcomes are related to different levels of the prognostic factor, this factor considered on its own. This is an often used first step to reduce the number of potential prognostic factors that have to be considered in future multivariable analysis, eliminating those factors that have no significant differences in outcome, considered in this way. This univariable screening process carries with it the risk of failing to identify an important covariate that will show itself important only after correcting for levels of another prognostic factor.\textsuperscript{60} In our multivariable analysis, no reduction of variables was done: even non-significant variables were repeatedly, at every forward step in multivariable analysis, evaluated for significant contribution. The only variable in our studies that turned out to be important after being non-significant in univariable analysis (step 0 in the Cox proportional hazards model) was the factor “pain” (Chapter IV).

Again, the evolution of the survival probability in time can be displayed as a survivor curve using the Kaplan-Meier method. At this stage, Kaplan-Meier analysis allows us to compare the survival probabilities of two (or more) different groups of patients at any given time in follow-up.

To test the null hypothesis that two (or more) survival curves are identical,
the most frequently used statistical test is the log-rank test. In a prognostic factor study with multiple prognostic factors analyzed, often a 1% significance level is more appropriate than the usual 5% level, and this is the level we used in our studies in Chapters II and III. The log-rank test is a statistic that provides an overall comparison of the Kaplan-Meier curves being compared, using the difference between observed and expected number of events in each group at each point in follow-up where a death occurred, the expected number of deaths being calculated from the assumption of no difference in survival. A P-value is most often derived from the chi-square distribution with one degree of freedom in case of comparison of two curves. If G groups and their respective survival curves are compared, there are G-1 degrees of freedom to derive the P-value from the X² value. An exact version also exists but was not used in this thesis.

At b.3. Multivariable analysis

Different prognostic factors can be highly related. Especially in the field of oncology, this is often the case. Examples are e.g. the observation that the larger the primary tumor is (T-status), the higher the likelihood of nodal metastasis (N-status) and subsequent irradical resection (involved margins). The older the patient, the more likely possible genetic damage, possibly resulting in a more undifferentiated and aggressive tumor. Especially in a situation like this, a multivariable analysis of the joined effect of different prognostic factors, is important.

If the effect of several potential outcome-influencing or outcome-associated factors is to be assessed in the same analysis, we can use:

- either a stratified analysis, assessing differences by the log rank test, and subsequently combining these into an overall evaluation. This can only be done in the case where a few categorical variables are considered.
- or a more flexible “multivariable analysis” (Cox proportional hazards analysis; recursive partitioning or tree analysis, neural networks, etc.)

The Cox model is based on the hazard function, and is very popular to analyze the joined effect of different prognostic factors on a specific patient outcome, because each prognostic factor can be studied after canceling out the confounding effects of other prognostic factors. Confounding occurs whenever the observed relationship between the levels of one candidate prognostic factor and the outcome is in fact due to another prognostic factor, who is strongly related to the outcome, being unevenly distributed over the levels of the first, candidate prognostic factor.

The outcome is duration of time to the occurrence of a binary “failure event” occurring during follow-up. The prognostic factor is assumed to affect the
risk of failure in a multiplicative fashion, that is constant over time, an assumption that needs verification upon using the Cox model. The model yields, for every variable under consideration, a significance test, a point estimate of its effect on the hazard rate, called the coefficient, corrected for associations with other covariates included in the analysis, and a confidence interval for this coefficient.

The coefficient gives the change in hazard rate per unit of the variable on the natural log scale. The hazard ratio per unit of the variable (and its confidence interval) is then obtained as an exponential of the coefficient (and its confidence interval), and this is the link between the results of the analysis and the interpretation in terms of compared survival functions.

It is interpreted as follows:

Example: gender as prognostic factor, women compared to men (men in the denominator)

HR = 1 => no relationship, no difference in survival function between men and women

HR < 1 => women relatively protected, hazard rate of women lower than that of men, survival function of women higher than survival function of men.

HR > 1 => men are relatively protected, hazard rate of women higher than that of men, survival function of women thus lower than survival function of men.

The P-value of the significance test for a potential explanatory variable is derived from the Wald statistic (obtained by dividing the coefficient for the variable by its standard error), that is assumed to have a standard normal distribution (Wald statistic = Z statistic). This Wald test is used in Chapter V, because of the computer program used (the SAS® package). A generally better but computationally more demanding alternative, the likelihood ratio (LR) statistic, is used in Chapter IV. Both statistics lead to a somewhat different P-value, but mostly lead to the same conclusions, except in certain extreme situations where the Wald test becomes unreliable. Fortunately such situations can be deduced from the results of the analysis. A third test based on the proportional hazards regression analysis, the score test, reduces to the log-rank test in the case of a single covariate. This test too is not provided within the SAS® package. There is generally no definite reason to choose between the LR test and the score test, while both are more reliable than the Wald test. In our case with a large data set and enough events and no overwhelmingly strong prognostic factors, all three tests give equivalent results.

The most frequent way to proceed in multivariable Cox proportional hazards analysis is called “stepwise regression”. The analysis consists of repeated assessment of all variables under consideration, progressively building a model containing only the factors that remain significant after adjustment for each other. In the type of stepwise analysis we have used in the study in
Chapter IV, at every step, the most significant factor that has not been selected yet, is added to the factors used to adjust for, if the accompanying $P$-value stays below a certain pre-specified value (in our study, the level $< 0.15$ was preset, with reassessment of significance for each selected factor in the model; a factor losing significance was to be removed once its $P$-value went $> 0.20$). Simon and Altman$^{73}$ and Marubini and Valsecchi$^{72}$ give an extensive overview of the caveats associated with the use of forward stepwise proportional hazards analysis, and also of possible alternatives.

As already mentioned under “quality checkpoints” (a.3), an interesting way to present the results from a multivariable analysis is the creation of a prognostic index. The aim is to combine several prognostic factors, each individually giving predictions with relatively low accuracy, into a single variable of high accuracy.$^{63,69,70}$ As in other fields of medicine, prognostic estimates based on clinical judgment alone tend to be far from correct, and several authors$^{67,68,73}$ feel that a properly validated prognostic index can lead to more accurate prognoses.

A prognostic index is constructed by summing the prognostic variables that remain important after multivariable analysis, each accompanied by a weight based on the regression coefficients as they emerged from the multivariable proportional hazards model. For every patient, subsequently a score can be computed, and this range of scores, incorporating the specific individualized information on prognosis, can then be subdivided to provide for “prognostic groups”, with the aim of being “homogeneous within” and “heterogeneous between”. This strategy is advocated by Simon and Altman: “It is preferable to produce plots based on prognostic categories defined by a prognostic index” (or to use a method to adjust the Kaplan-Meier plot for other variables).$^{73}$ Like other authors did before us in other fields of medicine (e.g. the Jass classification for rectal cancer), in the study in Chapter IV, we also constructed a range of scores, subsequently divided into four reasonably balanced groups with differing prognosis.$^{85}$

c. Biostatistical aspects: comparing two prognostic systems (Chapter V and Chapter VI)

A prognostic system is any combination of different individual prognostic factors, each individually giving predictions with relatively low accuracy, into a single variable of high accuracy.$^{63,69,70}$ Only very infrequently, for one specific disease, the predictive values of different prognostic systems are compared to one another.$^{66,86}$

c.1. Qualitative comparison.
Discrimination can be evaluated, by univariately considering the single vari-
able referring to the prognostic system, and constructing Kaplan-Meier curves.\textsuperscript{83} This allows for a visual idea of discriminative capacity and of preservation of correct ordering of patient groups. This way of evaluation was used in Chapter V (comparing the predictive ability of the prognostic indices in the source population versus the ability in the validation sample) and in Chapter VI (comparing predictive ability of the prognostic index PS1 to the UICC/AJCC stage grouping system). Other authors confine themselves to this analysis in comparing prognostic systems.\textsuperscript{87} However, this way of comparison remains rather subjective and is difficult in case of more than two groups.

c.2. Quantitative comparison

c.2.1. Other than merely visual assessment, the Cox proportional hazards model\textsuperscript{81} can be used to compare the predictive performance of one or more predictive systems in one or more data sets in a more quantitative way. The comparison of the predictive performance of one system in two different data sets was used in the external validation study in Chapter V. The difference in predictive performance of different prognostic systems in one data set was quantified in this way in Chapter VI.

c.2.2. Concordance measure C constitutes another way to make a quantitative comparison of prognostic abilities.\textsuperscript{72,86,88} It was initially developed for evaluating one prognostic system in two or more different data sets,\textsuperscript{89} in the situation where before starting analysis, the population is split into a “training” and “testing” population. The model is developed using only the “training” population, followed by a “validation” in the “testing sample”. Concordance measure C then allows for a quantitative comparison of prognostic performance in training and testing population. It was used in this way in Chapter V, where the population of the Dutch Head and Neck Oncology Cooperative Group was used as the “testing sample” to verify the prognostic ability of the prognostic index developed using the population of the Netherlands Cancer Institute (the “training” sample). It can be used in the same way to evaluate two different prognostic systems in one population, when both systems can be attributed a Concordance Measure C, that subsequently can be compared. This was the way we used it in Chapter VI.

Concordance measure C estimates the proportion of pairs of patients in whom a worse value of the prognostic system indeed implies a worse outcome. It corresponds to the probability that, for a randomly chosen pair of patients, predicted and observed outcomes are concordant.\textsuperscript{88} Analogous to the “Area under the ROC curve” for diagnostic tools, this measure can range between 0.5 (no discrimination ability, 50% of pairs of patients has a con-
cordance of predicted and observed outcome, but 50% have no concordance) to 1.0 (perfect concordance). It can be theorized, however, that the maximum concordance that can be reached while keeping adequately balanced and calibrated patient groups is 0.82 - 0.84.86,90,91

3. Prognostic research in malignant salivary gland neoplasms: where are we now, where should we go from here? Outline and aims of the study

3.1. Where are we now?

In addition to the variety of anatomic sites, the high number of different histologic subtypes and the different grades within these subtypes (Chapter I.1), many of the salivary gland tumor types have an indolent behavior and require long follow-up. These factors complicate the possibility for head and neck oncologists to accumulate experience and to evaluate treatment efficacy and with it of course the effect of modifying prognostic factors.4 Taking parotid carcinoma as an example, treatment results in major treatment centers are displayed in Table 5. These numbers have to be appreciated in their specific context of stage distribution, percentage of high grade tumors, treatment period and corresponding treatment regimens, and patient inclusion criteria.

<table>
<thead>
<tr>
<th>Authors</th>
<th>year of publication</th>
<th>DSS 5y</th>
<th>10y</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frankenthaler, et al.</td>
<td>1991</td>
<td>75%</td>
<td>70%</td>
</tr>
<tr>
<td>Kane, et al.</td>
<td>1991</td>
<td>69%</td>
<td>58%</td>
</tr>
<tr>
<td>Pedersen, et al.</td>
<td>1992</td>
<td>71%</td>
<td>65%</td>
</tr>
<tr>
<td>Poulsen, et al.</td>
<td>1992</td>
<td>78%</td>
<td>65%</td>
</tr>
<tr>
<td>Renehan</td>
<td>1999</td>
<td>55%</td>
<td>47%</td>
</tr>
<tr>
<td>Spiro</td>
<td>1993</td>
<td>77%</td>
<td>65%</td>
</tr>
<tr>
<td>Spiro and Wang</td>
<td>1989</td>
<td>63%</td>
<td>47%</td>
</tr>
<tr>
<td>Therkildsen, et al.</td>
<td>1998</td>
<td>76%</td>
<td>72%</td>
</tr>
</tbody>
</table>

For the patient groups and the corresponding treatment results within one center, prognosis for individual patients varies widely. Observing this, it is a
natural reflex for any health care professional working in this field of head and neck cancer, to try and define the relation between different outcomes reflecting treatment results and the variables defining these outcomes. To answer the question for this relationship, basically the tools as described in Chapter I.2 have to be applied to the study domain as depicted in Chapter I.1. Understandably, a large number of efforts to do this can be found in the literature of the last decades. Prognostic factors in salivary gland carcinoma have been studied extensively employing univariable and multivariable statistical analyses of patient and tumor characteristics in relation to treatment results. Histologic type and grade, stage, increasing age, gender, pain, skin invasion, and specifically for parotid carcinoma, facial nerve involvement (perineural growth), and treatment (resection margins) emerged as important prognosticators in this respect. However, due to differences in the study population of the reporting centers with regard to stage distribution, percentage high grade tumors, treatment period and corresponding treatment regimens, and patient inclusion criteria, it remains difficult to get a clear picture of what factor, to what a degree, is predictive of treatment results.

Most authors cite all these prognostic factors separately and expect the clinician to make an intuitive amalgam of an ever increasing number of factors, present in a given patient, to get an accurate estimate of prognosis. One way to deal with the increasing prognostic information is seen in the evolution from the 1992 staging system where factors such as facial nerve involvement and evidence of local extension contribute to a higher T classification. A number of studies investigating the added value of this changed classification do not offer very convincing arguments to believe that it is really an improvement. The question can be asked whether it is appropriate to mix the information of tumor size with other variables like facial nerve invasion, when these two aspects can occur independently and can be found independently prognostically important in several multivariable analyses.

3.2. Where should we go from here: aim and outline of this thesis

The studies compiled in this thesis were performed with two aims. The first was a descriptive one: to get an idea as accurate as possible of the patient population treated in our Institute in the past three decades. This information is described for submandibular carcinoma in Chapter II, for carcinoma of the minor salivary glands in Chapter III, and for parotid carcinoma in Chapter IV.
The next and more important aim was to reach, for the different anatomical locations considered, as high a level as possible on the "prognostic ladder", the hierarchy of prognostic research, as described in Chapter I.2, based on the information contained in the patient group treated in the Netherlands Cancer Institute. In this population, of reported prognostic factors, what are the relevant ones, and to what degree they are relevant?

Where possible, the aim is to go beyond the level of mere identification/confirmation and (eventual) quantification. The latter level is where most prognostic factor studies dealing with salivary gland carcinoma stop and leave the reader helpless with a list of factors and corresponding $P$-values, without a clear indication of how to apply this knowledge in clinical reality.

Several limitations of our study material have already been highlighted in Chapter I.2, when the ideal situation of prognostic research was summarized and thus, unfortunately, for the anatomical subgroups considered in Chapter II and III, the available number of patients and the events occurring limit the possibility of prognostic research to step 1 on the prognostic ladder. These patient groups have a detailed clinical record and a strong follow-up, however, the restriction to univariable analysis limits our prognostic research to a solid identification-confirmation of prognostic factors in our material.

In Chapter IV, the group of patients with parotid carcinoma allows us to step up one level on the prognostic ladder. Following univariable identification, multivariable analysis allows for both identification of factors and quantification of their respective effect. In stead of stopping at the level of reporting the different resulting factors and their respective weight in terms of a hazard ratio, we wanted to report our results in as practical a way as possible, and we therefore chose to create a prognostic index, combining the different independent prognostic factors remaining after multivariable analysis. This index sums the properly weighted contributions of each patient and tumor characteristic to prognosis into a single number, corresponding to an estimate of tumor recurrence following primary treatment.

Chapter V describes the next step on the prognostic ladder, adding scientific support to the prognostic index developed in Chapter IV. This next step was the testing of the index, created using the patient population of the Netherlands Cancer Institute, in an independent population. The aim here was to add further evidence to the usefulness of the devised prognostic index by indicating generalizability to geographically different patients. For this purpose, we had the opportunity to use the patients included in a nationwide database. Following this external validation, in a search for practical applicability, we constructed a downloadable easy-to-use computing program which allows the head and neck oncologist to easily obtain a quantitative estimate of prognosis tailored to the patient he is treating at that moment.
This is the highest level of prognostic evidence we have reached in this thesis. The next step up on the prognostic ladder, and thus an even increased level of credibility of our results, can be reached by multiple independent validations, as stated by Justice.\textsuperscript{77} It is of course also possible that we have to step down following an independent validation not confirming our findings. Progress in science can only be made by the continuous creation of rejectable hypotheses (Sir Karl E Popper, 1984), and the future will have to point out how long ours will stand firm.

In Chapter VI, we leave the prognostic ladder concept, and attempt to compare the prognostic ability of the created prognostic index to the “established” UICC-AJCC TNM staging concept. For this purpose we again used the nationwide database of the NWHHT, and applied the techniques as described under 2.2.3.

Chapter VII is a literature review on management of the largest subgroup of patients with salivary gland carcinoma, the patients with parotid carcinoma, integrating the information from this thesis in the current knowledge base on this topic.

In Chapter VIII, a general conclusion with an outlook to the future is formulated.
References


Chapter I


71. Marubini E, Valsecchi MG. Competing risks. Analysing Survival Data from Clinical Trials