clinical decision making in oncology

with special reference to patients with cancer of the head and neck

lukas stalpers
clinical decision making in oncology
with special reference to patients
with cancer of the head and neck
Contents

Part A: Clinical Decision Making in Oncology

Preface 1-2
1. Introduction: Clinical decision making and oncology 3-5
2. The structure of decisions: The decision tree 7-18
3. Evaluating outcomes: The adjusted DEALE 19-23
   (L.J.A. Stalpers, J.J.M. van Gasteren, W.A.J. Van Daal)
5. Multiple outcomes: Utility assessment 33-54
   5.1 Introduction: Cancer Treatment, Quality of Life and Decision Making
   5.2 The Life Standard Gamble
   5.3 Torrance’s Time Tradeoff Technique
   5.4 Additive Conjoint Measurement

Part B: Applications

6. Results of radiotherapy and surgery for glottic carcinoma 55-68
   (L.J.A. Stalpers, A.L.M. Verbeek, W.A.J. Van Daal)
7. Radiotherapy or surgery for T2N0M0 glottic carcinoma: A decision analytic approach 69-84
   (L.J.A. Stalpers, A.L.M. Verbeek, W.A.J. Van Daal)
8. The role of yearly chest radiography in the early detection of lung cancer following oral cancer 85-98
9. Yearly chest radiography in the early detection of lung cancer following laryngeal cancer 99-110
10. Markov modelling in clinical trial design 111-140
11. Utility assessment by time tradeoffs: A delusive deal 141-149
    (L.J.A. Stalpers, A.L.M. Verbeek, W.A.J. Van Daal)
12. Additive conjoint measurement: Applications in medical decision making 151-176
    (L.J.A. Stalpers, A.L.M. Verbeek, W.A.J. Van Daal)
Part C: Appendices:

C.1 The life standard gamble 177-178
C.2 The time tradeoff test 179-181
C.3 The additive conjoint measurement test 183-192
C.4 FORTRAN routine (+example) to identify circularities in a preference ranking from paired comparisons. 193-198

Samenvatting 199-200
Preface

Acknowledgement

This thesis presents the results of a study running from November 1985 to November 1988. This study was initiated and conducted by Prof. Dr. W.A.J. van Daal, head of the Institute for Radiotherapy and Dr. A.L.M. Verbeek from the Department of Epidemiology, both at the Catholic University of Nijmegen, and amply profited from the cooperation with Prof. Dr. Th.G.G. Bezembinder from the Nijmegen Institute for Cognition-research and Information-technology (NICI) and with Dr. J.H.M. Zwetsloot-Schonk from the Department of Clinical Oncology of the University of Leyden.

Data collection for the retrospective studies was performed with the help of A. Engelen, M.D., P. Van Vierzen, M.D., and the staff of the Departments of Diagnostic Radiology (Prof. Dr. J.H.J. Ruys), Maxillofacial Surgery (J.Brouns, M.D., D.M.D., and I. Bruaset, D.M.D.) and Otorhinolaryngology (Prof. Dr. P. van den Broek, Dr. J.J. Manni). For the analysis of the data from the additive conjoint measurements, I had much help from and fruitful discussions with A. Bendermacher, M.Sc., F. Gremmen, M.Sc. and A. Maas, M.Sc. from the NICI and C. Verhoef, M.D. and J. van Gasteren, Ph.D. from the Institute for Radiotherapy, Nijmegen. Most of the English corrections were performed by Mrs. C.W. McKell.

I am most indebted to my dearest colleague and beloved friend Patty Nelemans: I could always rely on her advice and love in the most desperate moments.

The structure of this thesis

Figure 1 graphically displays the structure of this thesis. Part A is meant as the main text of this thesis. It should be regarded as an introduction to the structural principles and problems of clinical decision making in oncology. Part B gives some extensions and applications from the field of diagnosis and treatment of patients with cancer of the head and neck. Part C gives the appendices to parts A and B. Since we were primarily concerned with the cure and care of patients already diagnosed with cancer, this thesis is focused on optimizing choices concerning
medical treatment in individual patients. The evaluation of diagnostic tests is not discussed in this thesis.

Figure 1 • Structure of this thesis
1. Introduction: Clinical decision making and oncology

Uncertainty is intrinsic to clinical practise in choosing and prescribing a certain treatment from several alternatives a physician is uncertain about the outcome in a patient, how the outcome is qualified by the patient or whether or not an alternative treatment might yield a better outcome [Sox et al., 1988]

Clinical decision making is a rational approach to reduce uncertainties in choosing the best among several strategies in clinical practise and in health care planning. It may thus provide more certainty about uncertainties. Clinical decision making unifies the common sense of clinical practise with the probabilistic approach of clinical epidemiology, psychological measurement techniques and management sciences. Although the origin of rational decision making at the patients’ bedside goes back to the early Hippocratic days of medicine a more formal approach comes from the field of philosophy, management science, operations research and gamble theory [Bayes, 1763, Castles et al., 1971, Krischer 1980, Luce and Raiffa, 1957, Raiffa, 1968, Scott, 1967, Von Neumann and Morgenstern, 1953 Weinstein and Fineberg, 1980]

Cancer is the collective noun for a group of diseases characterized by growth and spread of tissue beyond the normal anatomic boundaries which, untreated leads to physical, mental and social deterioration or death. The goals of clinical care of patients with cancer can be summarized by the classical commitments of the medical profession. Sometimes to cure frequently to relieve and always to comfort patients with cancer [Pare, 1510-1590] Since the commitments can be conflicting, both physician and patient have to consider the risks and outcomes of several treatments. Clinical decision analysis can be useful in structuring and evaluating such difficult choices.

Clinical decision analysis in oncology may be limited for three impediments concerning

1) The validity of risk estimates, or probabilities, for particular events, which can be imprecise or even unknown, especially considering the conditionality or multifactorial determination of certain events. For instance, The mortality rate of an individual patient with cancer of the vocal cords is not merely determined by the extent of the tumour and by the treatment received, but also by age, sex,
recurrence rates, the chance to be salvaged following recurrence, the risk of developing new malignancies and by a multitude of more or less defined other factors

(2) The validity of value judgements. The outcomes or 'health states' resulting from a medical treatment and hence the treatment itself, can only be judged by its subjective value, measuring the patient's value judgement of an outcome is called utility assessment. Methods for utility assessment are still in an early phase of development and have not yet been applied to clinical decision making.

(3) Time. Both risk estimates and value judgements are strongly dependent on time. Tumour recurrence rates and subsequently mortality rates from cancer generally decrease in time but death from other causes (accidents, heart disease, new cancers etc.) increases in time. Similarly, value judgements about the length of life and the quality of life may shift over time, as the data of this thesis suggest, one will generally attach more value to close life years than to distant life years. Consequently, the value (or utility) attached to a worse health state compared to the best possible health state, is inversely related to life expectancy. Treatment decisions in oncology have therefore to be evaluated considering the chronical aspects of the course of disease and the time-dependency of utilities.

Although these three factors are not specific for oncologic practise, they are characteristic for the problems encountered in decision making for individual patients with cancer and for health care planning in oncology. This thesis will therefore focus on these three issues. Chapters 2 and 3 are primarily focused on structuring and evaluating decision problems using fixed outcomes and static probabilities, chapter 4 gives the Markov model structuring decisions in a dynamic way to enable analysis with time varying probabilities and utilities, chapter 5 describes methods for utility assessment.

References


2 Castles FG, Murray DJ, Potter DC (Eds) *Decisions, organizations and society* Selected readings Penguin Education-The Open University, Harmondsworth (GB) 1971


5 Pare, A. Original text. *Guérir parfois, soulager souvant, consoler toujours.* The original source of this phrase, which is attributed to Ambroise Paré (1510-1590), has been lost. The phrase decorates the amphitheater of the Académie de Médecin in Paris (personal note Prof. B. Schulte).


8 Sox HC, Blatt MA, Higgins MC, Marton K. *Medical decision making.* Buttersworth, Boston 1988


2. The structure of decisions: The decision tree\(^1\)

2.1 The elements of a decision tree

The decision tree forms the main tool in structuring and evaluating the essentials of clinical decision making under uncertainty. The structural elements of clinical decision making and decision tree analysis are:

1. **Define the decision problem**
   What is the clinical problem faced by the doctor and patient? For instance: What is at this moment the best treatment for this specific patient with this specific type of cancer?

2. **Identify the decision alternatives**
   What treatments can be given to this patient, e.g. surgery, radiotherapy, chemotherapy or combination therapy?

3. **Structure the decision problem in time**
   What may happen following treatment: Operative death, cure, tumour recurrence, cancer death? In what order do these events occur during the course of the disease: Operative death is an event directly following surgery; cure can be obtained either directly following treatment or following treatment of a tumour recurrence.

4. **Define the outcomes**
   What are the final outcomes of each treatment and in what terms are they expressed? What is the decision criterion? Quantitatively: Survival rate, cure rate, life-expectancy. Qualitatively: Grade of mutilation, physical impairment and other more subjective aspects of the quality of life.

5. **Assign probabilities**
   What are the chances of the above events and outcomes expressed in probabilities or rates, e.g. death rate, recurrence rate, cure rate following recurrence?

6. **Assign utilities**
   What value judgement or utility should be attached to the respective outcomes, compared to the best and worst outcomes? For instance: What is the utility of 'cure with mutilation' and 'immediate death' compared to complete health?

---

Given the estimates of probabilities and utilities, the decision tree is used to calculate the expected effectiveness of each treatment. The expected effectiveness is expressed by one of several quantitative decision criteria. Frequently used criteria by which cancer treatments are compared are life-expectancy and survival rate. Life-expectancy is the average future lifetime of a patient at a specified age. Survival rate at a certain time $t$ is the fraction of patients surviving to time $t$. If the quality-of-life is considered in a comparison of treatments, utility theory can be used to weigh the life-expectancy following each treatment by the utilities of the respective outcomes. The treatment yielding the most quality-of-life adjusted life years (shortly QALYs), is commonly chosen as the preferred or 'best' treatment [Weinstein & Stasson, 1977, Williams, 1985]. Considering costs with utilities is denoted as cost-utility analysis [Drummond, 1977, Smith, 1987].

Since both the estimates of probabilities and utilities and the subsequent decisions harbour uncertainty due to insecurity of the values, a so-called sensitivity analysis is performed to check for the stability of a decision under varying estimates of observed or reasonably permitted values.

2.2 An example: A patient with cancer of the vocal cords

Problem definition
A 60-year old man with T3N0M0 glottic carcinoma is admitted for treatment. Should this patient be advised surgery or radiotherapy?

Structuring the decision problem
Figure 2.1 gives the decision tree structuring the choice between radiotherapy and surgery for this patient, including the recurrence rate following primary treatment and the salvage rate following recurrence. In patients with T3N0M0 glottic carcinoma, surgery usually consists of total laryngectomy. Laryngectomy can be performed both as primary treatment and as salvage treatment following failure to primary radiotherapy. Salvage treatment following failure to primary surgery generally consists of re-surgery with or without radiotherapy.
**Figure 2.1** - Decision tree for the choice between radiotherapy and surgery in a 60-year old man with a $T_3N_0M_0$ glottic carcinoma.

**LE:** Life-Expectancy (expressed in years)

**EU:** Expected Utility (expressed in QALY's)

**QALY:** Quality of life Adjusted Life Year

---

**Decision Tree**

- **Radiotherapy**
  - **Cure**: 0.58
  - **Salvage**: 0.37
  - **Recurrence**: 0.42

- **Surgery**
  - **Cure**: 0.73
  - **Salvage**: 0.26
  - **Recurrence**: 0.27

**Cases**:

- **$T_3N_0M_0$ Glottic Carcinoma**
  - **Radiotherapy**
    - **Alive, Natural Speech**: 13.69 years, 13.13 QALY's
    - **Alive, Artificial Speech**: 2 years, 0.53 QALY's
    - **Death from Cancer**: 2 years, 0.53 QALY's
  - **Surgery**
    - **Alive, Artificial Speech**: 14.73 years, 11.78 QALY's
    - **Death from Cancer**: 2 years, 0.32 QALY's
Following successful radiotherapy, a patient is able to speak with a natural speech. Following laryngectomy, a patient has to learn some form of artificial speech (oesophageal speech, speech by an electronical device or by a so-called Groninger button) and some 5% of patients will never recover their speech.

A *decision node*, denoted by a box, dissects the decision tree into the alternative treatments RADIOTHERAPY in the upper branch and SURGERY in the lower. The *outcomes* are given at the extreme right in boxes, being ALIVE, NATURAL SPEECH following cure by radiotherapy, ALIVE, ARTIFICIAL SPEECH following recurrence and salvage and DEATH FROM CANCER following treatment failure of a recurrence. The *chance nodes*, denoted by a circle, divide the course of disease into events to which the respective probabilities are attached. The recurrence rates and salvage rates following radiotherapy and surgery for T3N0M0 glottic carcinoma are mean estimates based on a literature study described in chapter B 1. The stream of events following surgery is similar to the events following radiotherapy but for two important exceptions:

(a) Following surgery, both recurrence rates and salvage rates are lower than following radiotherapy.
(b) All surgically treated patients have to learn an artificial speech, whereas successfully cured patients following radiotherapy are able to speak naturally.

As an index of comparison of both treatments, figure 2.1 gives both the life-expectancy (LE) expressed in years and the expected utility (EU) expressed in quality adjusted life years (QALYs).

**Evaluating outcomes: Life-Expectancy**

For a successfully treated 60-year old male patient with glottic carcinoma, we presume a life-expectancy similar to the age-specific life-expectancy in the male Dutch population, being 17.9 years, both following successful primary treatment and following successful salvage surgery [CBS, 1985]. Following unsuccessful treatment of a tumour recurrence, death from laryngeal cancer generally occurs within several months. However, since it might have taken some years before recurrence, patients with a fatal recurrence will have lived some years, say 2.0 years. The respective life-expectancies are attached as leaves to the branches of the decision tree.
Based on the probabilities leading to each outcome and the life-expectancy of each outcome, the contribution of each branch to the average life-expectancy following radiotherapy and surgery can be calculated. For instance: The branch ALIVE, NATURAL SPEECH contributes $0.58 \times 17.9$ years = 10.38 years to the average life-expectancy following radiotherapy. The branches ALIVE, ARTIFICIAL SPEECH and DEATH FROM CANCER contribute $0.42 \times 0.37 \times 17.9$ years = 2.78 years and $0.42 \times 0.63 \times 2$ years = 0.53 years respectively, resulting in an average life-expectancy of 13.69 years following radiotherapy.

Following surgery, the decision tree yields an average life-expectancy of 14.73 years. Based on the best life-expectancy alone, surgery will be preferred over radiotherapy in a 60-year old male patient with $T_3N_0M_0$ glottic carcinoma.

**Evaluating outcomes: Quality Adjusted Life Years**

In the previous analysis, the quality of life is not considered in the choice between radiotherapy and surgery.

Since the quality of speech constitutes a major criterion of treatment success in glottic cancer, it is natural to consider it as a decision criterion in a decision tree analysis.

The quality of speech can be expressed by a numerical value or utility. There are several methods to measure utilities for decision analytic purposes. Some of these methods will be discussed in chapter 5. For illustrative purposes only, we will here use the time tradeoff technique to measure utilities [Torrance, 1972]. For the time tradeoff technique one presumes that the utility of a certain outcome, e.g. an artificial speech, lies somewhere on a utility scale ranging from 0.0 for the worst outcome, i.e. immediate death, to 1.0 for the best outcome, i.e. natural speech. The utility of an artificial speech is measured by assessment of the maximum number of life years with an artificial speech a patient is willing to tradeoff to live with a normal speech. The utility of an artificial speech is the ratio resulting from dividing the number of years with a normal speech by the number of life years with an artificial speech. For instance, if a patient is willing to sacrifice 4 years of 20 years with an artificial speech to live (16 years) with normal speech (utility=1.0), then the utility of an artificial speech is said to be proportional to the respective life years, i.e. 16 years/20 years = 0.8. This utility can be used to adjust the life-expectancies following radiotherapy and surgery. For instance, following SURGERY, figure 2.1 gives a cure rate of 0.73 resulting in ALIVE, ARTIFICIAL SPEECH with 17.9 life
The contribution to the life-expectancy is $0.73 \times 17.9$ years = 13.07 years with an ARTIFICIAL SPEECH. Multiplication of these life years by the utility of an ARTIFICIAL SPEECH yields the contribution to the expected utility (EU) of surgery expressed in QALYs: $0.8 \times 13.07$ years = 10.45 QALYs.

Analogously, we can weigh the number of life years for each outcome by its utility. 'Cure' following RADIOTHERAPY results in life with a NATURAL SPEECH contributing $0.58 \times 17.9$ years $\times 1.0 = 10.38$ QALYs to the expected utility of radiotherapy. 'Salvage' laryngectomy following a 'recurrence' after RADIOTHERAPY results in an ARTIFICIAL SPEECH (utility = 0.8) contributing $0.42 \times 0.37 \times 17.9$ years $\times 0.8 = 2.22$ QALYs to the expected utility of radiotherapy. For a patient who dies from cancer after a recurrence with 'no salvage', we presume that he will have lived 2 years with a normal speech, hence resulting in a contribution of $0.42 \times 0.63 \times 2$ years $\times 1.0 = 0.53$ QALYs to the expected utility of radiotherapy.

The expected utilities (EU) following radiotherapy and surgery are 13.13 and 11.78 QALYs respectively. Under the main principle of utility theory, namely utility maximization, the treatment with the highest expected utility should be preferred [Von Neuman and Morgenstern, 1953] So, going by QALYs, in the example in figure 2.1 radiotherapy should be preferred to surgery.

Sensitivity Analysis: The stability of conclusions

The calculations of both life-expectancies and expected utilities are only valid under the given or 'baseline' probabilities and utilities. For instance, (1) What would happen if the utility of an artificial speech was not 0.8 but much higher (i.e. what if a patient was willing to sacrifice less than 20% of his life-expectancy to live a life with natural speech instead of with an artificial speech)? And (2) what if the recurrence rate following radiotherapy was higher, i.e. 0.50 instead of 0.42?

(1) Figure 2.2 gives the results of the sensitivity analysis on the expected utilities of radiotherapy and surgery for varying utilities of life with an artificial speech. Figure 2.2 demonstrates that, for a utility of artificial speech ranging from 0.5 to 1.0, the expected utilities of both radiotherapy and surgery increase. However, as appears from a steeper curve for surgery than for radiotherapy, the expected utility of surgery is more sensitive to variation of the utility of artificial speech than that of radiotherapy. The curves have a break-even point for which the expected utilities of radiotherapy and surgery are equal. The utility corresponding with the break-even
The decision tree

point (utility = 0.91) is called the *threshold utility*: Based on the expected utility, a patient who attaches less than the threshold utility of 0.91 to life with an artificial speech will be advised radiotherapy, a patient with higher utility values will be advised surgery.

Figure 2.2 - Expected utility of radiotherapy and surgery expressed in quality adjusted life years (QALYs) after radiotherapy and after surgery for a patient with a T3N0M0 glottic carcinoma for varying utilities attached to the quality of voice after laryngectomy (artificial speech). The utility of an artificial speech is expressed on a utility scale ranging from 0.0 for death to 1.0 for natural speech.

(2) Figure 2.3 gives the results of a sensitivity analysis on the threshold utility of an artificial speech for varying recurrence rates following radiotherapy. For a recurrence rate ranging from 0.3 to 0.7, the threshold utility of an artificial speech will decrease from 1.0 to 0.65. For pairs (utility, recurrence rate) below the threshold curve, the expected utility of radiotherapy is higher than that of surgery. From this figure we can read that for the baseline recurrence rate of 0.42 following radiotherapy, surgery is to be preferred if the utility attached to an artificial speech is higher than 0.91. However, if we presume a much higher recurrence rate of 0.50 following radiotherapy, we can read from figure 2.3 that a lower utility can be
accepted to have surgery preferred, i.e. utility > 0.84. However, if the utility of an artificial speech is 0.80, as we earlier presumed, radiotherapy will still be preferred, even for such a high recurrence rate. In the latter case, figure 2.3 shows that surgery is only to be preferred if the recurrence rate following radiotherapy ranges above 0.54. Here we always presume, of course, that the other probabilities (the salvage rate both following radiotherapy and surgery and the recurrence rate following surgery) remain stable.

Figure 2.3 • Threshold utilities for the quality of speech following laryngectomy (artificial speech) on varying recurrence rates for T₃N₀M₀ glottic carcinoma.

What if both the recurrence rates following radiotherapy and surgery are higher than the baseline values assumed?

Figure 2.4 gives a two-way threshold analysis on the recurrence rates following radiotherapy and surgery for the respective utilities of artificial speech 0.7, 0.8, 0.9 and 1.0. It should be recognized that the analysis using a utility of artificial speech of 1.0, which by definition is equal to the utility of natural speech, is in fact a two-way sensitivity analysis on life-expectancy. For instance, if the recurrence rate following radiotherapy is 0.50 and the recurrence rate following surgery is 0.40, then radiotherapy will only be preferred if the utility of artificial speech is only slightly less than the maximum utility of 1.0.
Based on the utility theory, a decision tree analysis offers a simple tool to structure and formally evaluate decision problems. It may be clear that considering the uncertainty of the probabilities and utilities, a decision tree analysis will not generate a clear cut advice for a specific treatment. The goal of decision tree analysis is to indicate within what range of values the decision problem is at stake and when it is relevant to consider other decision criteria rather than life-expectancy alone, for instance the quality of speech, and to what extent the assessment may be pivotal in the decision at stake. The ultimate choice of a treatment predominantly depends on the credentials the doctor and patient give to the most uncertain factors of the analysis, namely the validity of both probabilities and utilities.

Figure 2.4 • Sensitivity analysis of the threshold utility by varying the recurrence rate following radiotherapy, and/or the recurrence rate following surgery and for stepwise variations of the utility of an artificial speech from 0.7 to 1.0.
2.3 Probability Assessment

Several sources can be used to assess probabilities for decision tree analysis. Most commonly used sources are expert opinions, published data and data bases.

**Expert opinions**
Expert opinions and other sources that draw on personal experiences belong to the most hazardous forms of probability assessment. Many studies have shown the pitfalls in probability assessment due to several biases, amongst which selective memory and wishful thinking [Kahneman et al., 1982; Spetzler & Stael von Holstein, 1975; Tversky & Kahneman, 1974]. It is therefore advised to use probabilities assigned by experts only as an ultimum refugium in assigning risks and rates, or as indicators for maximal or minimal tolerable risks.

**Published data**
Data published in medical journals and books, -shortly denoted by 'the literature'-, may provide reasonably reliable estimates of probabilities. There are three limitations to the use of published data for utility estimates:

The first drawback is concisiveness. Since most journals prefer short and concise papers to extensive and detailed ones, only the most general rates will be published such as survival and recurrence rates in their most concise form. In a literature study of the results of radiotherapy and surgery in glottic carcinoma described in chapter B.1, detailed rates such as the salvage rate following tumour recurrence were only scarcely and imprecisely described.

Secondly, for the same reason of concisiveness, survival rates and recurrence rates are more often than not represented by only one endpoint such as the five-years rate, without considering the temporal characteristics of the course of disease. Such one-point measurements are valid only for events which occur only once at a specific time, such as operation mortality. As described earlier the course of disease in cancer is generally better characterized by a sequence of risky events, for which the probabilities change with time. A description of rates with preservation of the time-dependent characteristics may give a more accurate insight in the course of the disease. For instance, in the literature it is said that patients with oral cancer have an enhanced risk of developing lung cancer warranting regular screening for lung cancer [De Vries, 1987]. From our own study, described in
chapter B 3, the incidence of lung cancer following oral carcinoma was both calculated as a crude rate, being 22/213 = 0.103 and as an actuarial incidence rate, taking into account the time of clinical follow-up. The latter description yielded a two-years cumulative incidence of 0.13 and showed no further rise after these two years. Since no new lung cancer was detected after two years, it seems as if there is no more enhanced risk for lung cancer and that intensive screening for lung cancer becomes superfluous after two years follow-up.

A third hazard of published papers is the inclination of medical journals to publish positive results and to publish negative or less favourable results only in a later phase. Sackett gives an example of the evolution of the value of the carcino-embryonic antigen (CEA) in the early detection of colon cancer [Sackett, 1982]. The CEA test was originally introduced in 1969 as a test with a high diagnostic sensitivity being 97% (the percentage of colon cancer patients who had a positive CEA test result). However, in subsequent studies performed in less selected populations the sensitivity dropped to 90% in 1970, 87% in 1971, 82% in early 1972 and 72% later in 1972. A similar decline was seen for the diagnostic specificity (the percentage of patients without colon cancer who had negative CEA test results).

Broad and Wade [1982 & 1985] have described several cases of malpractice to publishing faked results, some scientists yield to temptation to commit fraud, forced by the preference of journals to publish positive results, in combination with the need of research grants that are awarded based on number of publications.

Data bases

Data bases, such as the classical medical files and more recently the hospital computer systems, are among the most reliable sources to assess probabilities, although they store a relatively small number of relevant data for specific (hospital) populations. The usefulness of data bases is highest for prospective studies, i.e., in which the questions which have to be solved and the relevant variables are structured in advance. For retrospective studies (in which the questions of the study are posed following the collection of data), relevant information is frequently missing or inaccurately recorded in the medical files. Compared with expert opinions and the concise data from the literature, data bases allow us to provide more accurate answers on more specific and more complex questions. It is our conviction that, wherever possible, a prospective clinical trial should be preceded.
by a retrospective study of own data to simulate and analyse the trial with the most appropriate data. Based on a retrospective study in patients with laryngeal cancer described in chapter B.4, chapter B.5 shows how these data were structured and analysed to give accurate estimates of the effect of bronchoscopy for the early detection of secondary lung cancer, providing some indications in the design of a randomized controlled trial.

2.4 References

3. Evaluating outcomes: DEALE-ing with Life Expectancy and Mortality Rates

3.1 Introduction

In 1982, BECK et al. presented a simple method known as the "DEALE" method, to approximate the life expectancy for an individual patient using information from various sources, such as disease specific survival rates and age- and sex-specific life expectancies from a table of vital statistics [Beck et al., 1982a]. In a subsequent paper, the use of the DEALE method in quantitative decision making is demonstrated [Beck et al., 1982b]. The DEALE is convenient to estimate life expectancy for a patient with particular disease characteristics. However, as we demonstrate in this report, the annual mortality force used in the DEALE formula is not similar to the annual mortality rate, despite the similar unity "per year". Therefore, the DEALE method is not appropriate and can actually be rather confusing in dealing with patients' questions about the average risk of dying next year or the chance of survival the next five years. In a decision analytical context, these questions are of particular importance for the assessment of both the probabilities used in a simple decision tree analysis and the transition probabilities used in a Markov model [Beck and Pauker, 1983; Beck, 1988]. In this short communication, we propose a slightly adjusted version of the DEALE appropriate to estimate both life expectancy and the instantaneous mortality rate.

3.2 The DEALE

The DEALE, which stands for 'decreasing exponential approximation of life expectancy' is based on the assumption that the survival curve \( S(t) \) follows a simple declining exponential function of time, assuming an annual mortality force \( \mu \) that is constant for all following times \( t \):

\[
S(t) = e^{-\mu t}
\]  \{1a\}

---

1 Lukas J.A. Stalpers, M.D., Hans J.M. van Gasteren, Ph.D. and Willem A.J. van Daal, M.D., Ph.D. Published in Medical Decision Making, 1989; 9: 150-152
The annual mortality force $\mu$ can be inferred by a logarithmic transformation of $S(t)$

$$\mu = -\frac{\ln S(t)}{t} \text{ per year} \quad \{2a\}$$

The life expectancy $LE$ is the inverse of the mortality force $\mu$:

$$LE = \frac{1}{\mu} = -\frac{t}{\ln S(t)} \text{ year} \quad \{3a\}$$

By combining information from different sources, the DEALE can be used to estimate the life expectancy of an individual patient, based on patient-specific characteristics such as age, sex, stage of disease and concomitant diseases. Using mortality forces $\mu_1, \mu_2, ..., \mu_k$ for $k$ prognostic factors, DEALE approximates survival rate, mortality force and life expectancy as follows:

$$S(t) = e^{-(\mu_1 + \mu_2 + ... \mu_k)t} \quad \{1b\}$$

$$\mu = \mu_1 + \mu_2 + ... \mu_k \text{ per year} \quad \{2b\}$$

$$LE = \frac{1}{(\mu_1 + \mu_2 + ... \mu_k)} \text{ years} \quad \{3b\}$$

For example: The life expectancy of a 60-year old male patient with a $T_2N_0$ glottic carcinoma treated by radiotherapy can be calculated from (1) the disease-specific survival of all patients with $T_2N_0$ glottic carcinoma treated by radiotherapy, and (2) the age-specific life expectancy of a 60-year old male.

(1) A recent review of 26 articles on results of treatment of glottic cancer gives an average five-year disease-specific survival rate of 70% for patients with $T_2N_0$ glottic carcinoma [Stalpers et al., 1987]. The DEALE yields a disease specific mortality force $\mu_D$:

$$\mu_D = -\frac{\ln 0.7}{5} = 0.071 \text{ per year}$$
(2) A table of vital statistics gives a life expectancy of 17.9 years for males aged 60 [CBS, 1987]. The DEALE yields an average age- and sex-specific mortality force for the period beyond 60 years $\mu_{AS}$.

\[ \mu_{AS} = \frac{1}{17.9} = 0.056 \text{ per year} \]

The overall or 'compound' mortality force $\mu$ for this specific patient is $\mu_D + \mu_{AS} = 0.071 + 0.056 = 0.127$ and hence the life expectancy $LE = 1/0.127 = 7.87$ years.

3.3 The 'adjusted' DEALE

The mortality force $\mu$ does not represent some readily interpretable rate or probability estimate. We therefore propose an adjusted version of the DEALE, based on the instantaneous mortality rate $\beta$, to predict survival rates, the instantaneous mortality rate per time-interval and life expectancy. The adjusted DEALE is based on the assumption that survival follows an exponential function on the complement of the instantaneous mortality rate $\beta$:

\[ S(t) = (1-\beta)^t \quad \{4a\} \]

with the instantaneous or annual mortality rate $\beta$:

\[ \beta = 1 - e^{-\frac{\ln S(t)}{t}} \text{ per year} \quad \{5a\} \]

and the life expectancy $LE$:

\[ LE = -\frac{1}{\ln(1-\beta)} \text{ years} \quad \{6a\} \]
Although the mortality force $\mu$ bears the unity "per year", $\mu$ is obviously not an instantaneous or annual mortality rate, but a transformation of the instantaneous mortality rate $\beta$

\begin{align*}
\mu &= -\ln(1-\beta) \quad \{7a\} \\
\beta &= 1 - e^{-\mu} \quad \text{per year} \quad \{7b\} \\
&= 1 - e^{\frac{1}{\text{LE}}} \quad \text{per year} \quad \{7c\}
\end{align*}

For instance: The mortality force $\mu = 0.127$ for the patient with glottic carcinoma does not mean that the mortality rate in the next years averages 0.127 per year but

$$\beta = 1 - e^{-0.127} = 0.119, \text{ or } 11.9\% \text{ per year}$$

For $k$ prognostic factors, the adjusted DEALE gives

\begin{align*}
S(t) &= [(1-\beta_1)(1-\beta_2)\ldots(1-\beta_k)]^t \quad \{4b\} \\
\beta &= 1 - [(1-\beta_1)(1-\beta_2)\ldots(1-\beta_k)] \quad \text{per year} \quad \{5b\} \\
\text{LE} &= \frac{t}{\ln S(t)} = \frac{1}{\ln[(1-\beta_1)(1-\beta_2)\ldots(1-\beta_k)]} \quad \text{years} \quad \{6b\}
\end{align*}

For example: For a male aged 60 with $T_2N_0M_0$ glottic carcinoma, the adjusted DEALE may generate both the predicted survival rate, an instantaneous mortality rate per year and a life expectancy in years:

1. Based on a five-year disease free survival of 0.7, equation $\{5a\}$ yields an instantaneous mortality rate:

$$\beta_D = 1 - e^{-t} = 1 - 0.931 = 0.069, \text{ or } 6.9\% \text{ per year}$$

2. Based on an age- and sex-specific life expectancy of 17.9 years, equation $\{7c\}$ yields an instantaneous mortality rate:

$$\beta_{AS} = 1 - e^{\frac{1}{\text{LE}}} = 0.054, \text{ or } 5.4\% \text{ per year}$$
(3) Equation (5b) yields the compound mortality rate per year:

\[ \beta = 1 - [(1 - \beta_D)(1 - \beta_{AS})] = 1 - (0.931)(0.946) = 0.119, \text{ or } 11.9\% \text{ per year} \]

(4) Equations (4a) and (6a) yield the expected five year survival rate.

\[ S(5) = (1 - 0.119)^5 = 0.53, \text{ or } 53\% \]

and a life expectancy:

\[ LE = \frac{t}{-\ln S(t)} = 7.87 \text{ years} \]

3.4 Conclusion

The original DEALE is convenient to estimate life expectancies using a compound of several prognostic factors, but it is less appropriate to calculate a compound mortality rate. The adjusted DEALE yields both a life expectancy and a more readily interpretable average mortality rate for a certain time interval, which may be used as an estimation of a probability in a decision tree or a transition probability in a Markov model.

3.5 References

4. Evaluating outcomes: The Markov model

4.1 Introduction

A simple decision tree and the DEALE-method are both used to describe the distribution of outcomes of a certain medical strategy for a fixed endpoint in time, e.g. to estimate a five-year survival rate or a life expectancy. The simple decision tree approach is most useful in clinical situations where events occur only once in a lifetime, mostly at the beginning of the course of a disease. Clinical events with such a unique and early occurrence are for example the risk of operative death and other acute complications of disease and treatment. The DEALE-method is most appropriate under the assumption that mortality rates are constant for all years following treatment. However, many clinical events in oncology are characterized by a chronic occurrence. Tumour recurrence, tumour metastasis, complications of treatment or cancer and subsequent death can occur at any time following treatment. The corresponding event rates will vary in time, in general, cancer related event rates decrease with time.

Another important time-dependent feature in generally older patients with cancer, is the steadily increasing age-specific mortality rate from other diseases. Chronic events may lead to a changing distribution of outcomes or health states over time and to variable durations spent in the subsequent health states.

Events that repeatedly occur in the course of a disease can be structured in a decision tree. However, if we want to model the course of a disease for about 20 years follow-up in a decision tree, defining a new decision tree for each year of the follow-up, the tree will rapidly grow into an inextricable brushwood with innumerable branches. For example, the simple decision tree described in chapter B 7, using three outcomes, namely NORMAL speech, ARTIFICIAL speech and DEAD-, will have more than 1000 branches following 20 years after primary treatment of glottic cancer. Analogously, the slightly more complex tree described in chapters B 10 and B 12 will have over one billion branches if modelled in a static decision tree.

Dynamic probabilistic systems can be used to structure repeatedly occurring chance events to generate more accurate estimates of the distribution of health states at any time and of the time spent in the respective states [Elashoff, 1984,  

---

1 A A Markov, 1856-1922 Russian mathematician [Ondar, 1981]
Sonnenberg, 1984, Mau & Steinke, 1986, Myers et al., 1980] Merely for its relative simplicity, the Markov model can be used as a convenient method for decision analytic purposes [Beck & Pauker, 1983] Paragraphs 4.2 and 4.3 discuss some specific problems in applying the Markov model to oncology. Some elaborate examples of the Markov model are described in chapters B.10 and B.12.

4.2 Principles and problems of Markov models

In a Markov model, the course of a disease is modelled in terms of probabilistic transitions from one state to another state in discrete time steps or 'Markov cycles.' The Markov process of prognosis is a finite probabilistic system [Kemeny & Snell, 1976].

In a Markov chain, the transition probability from one state to another is equal for every cycle [Jain, 1986, Silverstein et al., 1988]. Only a few examples of Markov chains in oncology have been described [Elashoff, 1984, Sonnenberg, 1984, Roach, 1988]. Indeed, it is very unlikely that relevant risks, such as recurrence rate and mortality rate, remain stable over time. In general, disease-related risks gradually decrease following treatment whereas age-related mortality risks increase over time.

In a Markov process, using time-dependent probabilities, the transition probability depends on the time passed since initiation of the process [see e.g. Myers et al., 1980, Mau & Steinke, 1986]. The Markov process is based on a major assumption, referred to as the Markov property [Kemeny & Snell, 1976]. The Markov property assumes that the probability distribution of health states after \( n \) time steps \( S_n \) only depends on the former distribution of health states \( S_{n-1} \). The associated transition probability \( P[S_n|S_{n-1}] \) neglects information about previous states. For this reason the Markov process is said to have 'a lack of memory' in determining the transition probability at each time step \( n \).

In chapter B.10, a three-state Markov process (WELL, LUNG CANCER and DEAD) is used to model the incidence and mortality from lung cancer in patients with laryngeal cancer with respect to the overall mortality. A possible course of disease is that of a patient who gets lung cancer at time \( = 1 \) and dies in one of the subsequent times, denote by time \( = j \).
The Markov model

<table>
<thead>
<tr>
<th>Time</th>
<th>0</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>WELL</td>
<td>→</td>
<td>LUNG CANCER</td>
<td>→</td>
</tr>
</tbody>
</table>

The Markov property assumes that the transition probability from WELL to LUNG CANCER at time $T = 1$ only depends on the previous health state at $T = 1-1$, and does not depend on the sequence of the preceding states. A possible violation of the Markov property is bypassed by presuming (1) unidirectional transition and (2) no other states in between than WELL. For unidirectional transitions, a hierarchical structure of transition states is presumed. Transitions 'down' from WELL to LUNG CANCER are allowed, but transitions 'up' from LUNG CANCER to WELL are not allowed. Since WELL is the initial state at time $T = 0$, the transition probability $P_1$ from WELL to LUNG CANCER only depends on $T = 1-1$. Since the DEAD state is the 'lowest' Markov state, and all patients die eventually, the DEAD state is called the absorbing state and the Markov model becomes a finite probabilistic system.

The Markov property further assumes that the transition Markov probability $P_j$ from LUNG CANCER to DEAD is independent from the time spent in the state LUNG CANCER. This is unlikely for patients with lung cancer. In general, the lung cancer survival curve strongly descends in the first 2-3 years following diagnosis and becomes constant after this period.

The modeling problem can be bypassed by several solutions. The first and most simple solution would be to presume a Markov chain, using a mean approximation of the mortality rate inferred from one or several points in the survival curve, e.g., from the five-years survival rate. The adjusted DEALE is a convenient method to infer such an average mortality rate. In the case of patients with lung cancer, this solution could falsely lead to an underestimation of the mortality rate in the near years and to an overestimation of the mortality rate for distant years.

A second solution would be to presume unidirectional transitions from LUNG CANCER to DEAD and to model a second Markov tree dependent on the time lung cancer occurred. However, such a second or higher degree in the Markov process is very cumbersome, with respect to modeling and evaluating.

A third solution is a quasi second degree Markov chain, introducing intermediate Markov states [Kemeny & Snell, 1976]. In chapter B 10, four intermediate LUNG CANCER states were defined, indicated by LUNG CANCER_2nd to LUNG...
CANCER_5th denoting the lung cancer states in year 2 to 5 following the occurrence of lung cancer. Presuming unidirectional transitions, the mortality probability $P_j$ (and its complementary survival probability $Q_j = 1 - P_j$) from lung cancer can then be modelled in a time-dependent fashion for the first five years following lung cancer. As the mortality rate from lung cancer has virtually become zero after five years, the number of intermediate states can be limited to five. After the last LUNG CANCER state, the process is reverted to the initial state WELL.

<table>
<thead>
<tr>
<th>Year</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>pNN</td>
<td>N</td>
<td>pNN</td>
<td>pNO</td>
<td>pOD</td>
</tr>
<tr>
<td>N</td>
<td>pNN</td>
<td>N</td>
<td>pNO</td>
<td>O</td>
<td>D</td>
</tr>
<tr>
<td>N</td>
<td>pNO</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>D</td>
</tr>
<tr>
<td>N</td>
<td>pNO</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>D</td>
</tr>
</tbody>
</table>

Scheme 4 1: All possible year-by-year courses of disease in patients with normal speech (N) after radiotherapy, followed by a laryngectomy after tumour recurrence, resulting in oesophageal speech (O) in the third, second or first year, and finally death (D) in the fourth year.

Example: In chapter B 12 a six-state Markov process is described to model the follow-up of patients with radiotherapy for glottic carcinoma. Scheme 4 1 shows all possible courses of disease in a hypothetical patient with a normal speech (N) after radiotherapy, followed by a laryngectomy after tumour recurrence, resulting in oesophageal speech (O) in the third, second or first year, and finally death (D) in the fourth year. The transition probabilities denoted by pNN, pNO, pOO and pOD (a fifth transition probability, namely pNO, is not used in scheme 4 1). If the transition probabilities are equal for all years, the Markov model is called a Markov chain; if the transition probabilities vary with each year, the Markov model is called a Markov process.
The Markov property assumes that the risk of dying with oesophageal speech in the fourth year following radiotherapy is equal for all laryngectomized patients, irrespective of whether a patient was laryngectomized for tumour recurrence in the first, second or third year following radiotherapy. This is unlikely as the mortality rate in the first year following laryngectomy is higher than in the subsequent years. In structuring the Markov model, this violation of the Markov property has been partially bypassed by presuming that the mortality from tumour following a recurrence is only increased in the year of recurrence and that the mortality from tumour in subsequent years is equal to the recurrence rate.

We do not agree with Beck that the Markov model is not fit for oncology '...while cancer clearly does not...obey the Markov property. In malignancies, the prognosis depends on which specific remission or relapse is being experienced' [Beck, 1988]. On the contrary, as demonstrated in chapters B.10 and B.12, the Markov process can be considered as a simple and powerful tool to model the course of a disease of a patient with cancer and to make accurate predictions about the prognosis of an individual patient and to estimate the survival in a group of similar patients.

4.3 Modelling time-dependent probabilities

Necessarily, actuarial incidence and mortality rates are only probabilistic estimates. Both for practical convenience and for a systematic analysis, combining and comparison of these rates, some mathematical model has to be assumed to frame the actuarial data. The choice of a mathematical model depends on three, -not necessarily exclusive or conflicting-, criteria:

1. Biological soundness: Model according to what you think is reasonable considering the theoretical concepts or assumed biological principles.
2. Experimental faithfulness: Model according to your data (i.e. according to some 'best fit' criterion).
3. Simplicity: Model what is practically most convenient.

For instance, figure 4.1 gives three mathematical approximations of the same actuarial estimates of the tumour recurrence-rate from a follow-up study in 556 patients treated for laryngeal cancer [B.9, B.10]. In order to describe a biological process like tumour recurrence, we may wish to model the data as a continuous function of time, in accord with one of three above mentioned criteria:
Figure 4.1 • The recurrence-rate from laryngeal cancer modelled according to:
(1a) A biological sound exponentially decreasing function (correlation = 0.75),

(1b) A best fitting criterion using a complex fifth order polynomial function (correlation = 0.99) and

(1c) A simple linear decreasing function (correlation = 0.90).
1. Biological soundness: A widely accepted theory assumes that tumour recurrence is highest just after treatment and gradually decreases during the following years, but that the threat of a recurrence will never completely cease [Dippel & Segaar, personal comment]. A mathematical approximation which meets this theory is an exponentially decreasing function of time $t$ with incidence accelerator $\lambda$:

$$\text{Incidence-rate} = e^{-\lambda t} \quad \{1\}$$

as drawn in figure 1a.

2. Experimental faithfulness: Figure 1b gives a complex mathematical function according to a 'best fit' approach, which (almost exactly) describes the experimental data.

3. Simplicity: According to the third approach, figure 1c gives a simple linearly decreasing function with the best data-fit.

The major drawback of the biologically sound approach is that it may result in wishful and conservative thinking. From an idealistic point of view, an over-reliance on theoretical concepts and an underestimation of the experimental data, may stray from new insights or new hypotheses about the 'real' structure of life.

The opposite can be said of relying too much upon the experimental data, since the data may be insecure due to statistical variation or observation biases, and may be (and probably are) determined by the interaction between several simultaneously acting biological processes which can only be fit by a complex mathematical model.

The drawback of the third approach is over-simplicity, hence violating both the most basic biological concepts about real life (e.g. a recurrence-rate cannot be less than zero, as suggested by figure 1c) or that it may poorly fit into the experimental data.

Although the most complete or ideal deterministic model may eventually be the mathematical model that optimally describes and predicts our data [Hawking, 1988], it is important to realize that in modelling for practical purposes, -such as in clinical decision analysis-, we have to be content with a compromise between the wishful, the experienced and the feasible.
4.4 References

5. Utility Assessment

5.1 Cancer Treatment, Quality of Life and Decision Making

Survival rates, cure rates and life-expectancy are major outcomes of treatment of patients with cancer. In recent years some powerful and sophisticated methods such as the Markov model for medical prognosis have been developed to generate accurate estimates of such quantitative outcomes, especially useful to medical decision making in oncology.

Aware of the fact that the preservation of the quality of life constitutes an important goal of cancer patient treatment, few physicians with experience in oncology will cling to survival and cure rates as the only criteria in choosing between alternative cancer treatments. However, since the quality of life seems to be a too weakly defined 'soft' and subjective matter, many physicians will hold to survival and cure rates as the main criteria of treatment success. A further reason for the scarce use of the concept of quality of life in medical practice and clinical decision making, can be attributed to the inexperience with and lack of convenient methods to measure the quality of life or relevant and decisive aspects of the quality of life.

The semantic and ethical issues of quality of life in medical practice have been extensively discussed by others [Mulder, 1983, Musschenga, 1987, Verkes, 1987, De Haes, 1988]. The present study deals with the practical problems of quality of life measurement for medical decision making in oncology.

In 1948 Karnofsky et al. introduced a bedside scaling method to give a quick and quantitative description of the general health-state of a patient with cancer [Karnofsky et al., 1948, Karnofsky and Burchenal, 1949]. The Karnofsky Performance Status Scale has become a popular tool to register the quality of life in cancer patients, especially for the evaluation of cancer clinical trials [Mor et al., 1984].

Many adaptations and variations have been developed based on the Karnofsky scale, ranging from simple bedside tools to elaborate questionnaires in order to obtain an impression of a patient's general health-state or quality of life.

---

1 Presented at the Second Conference of the European Society for Medical Decision Making, Copenhagen, June 1st 1988 [Stalpers et al., 1988]
[Van Dam et al., 1984] De Haes and Van Knippenberg give a detailed methodological review of the current methods to assess the general quality of life in patients with cancer [De Haes and Van Knippenberg, 1985] Most of these methods generate a more or less accurate general description of a patient's well-being. Such merely descriptive methods can be very convenient to monitor qualitative changes in the course of a disease in cancer clinical trials. However, such predominantly qualitative descriptions do not, by themselves, provide any norms that may guide medical decision making. Norms for quality of life are highly desirable to weigh against other attributes such as survival-rate, cure-rate and/or life-expectancy. The quantitative weighing of various qualitative properties on a common scale is called utility assessment. The utility of a certain attribute, for instance the quality of life-, is a value judgement (by patient and/or doctor) of the quality of life, and this is to be balanced against the utility of another attribute, for instance life-expectancy, generally assessed by a patient or a physician. The utility approach to medical decision making is based on the expected utility theory as described by Von Neuman and Morgenstern (1953). The expected utility theory states that individuals should assign (or act as if they assign) a utility to each possible outcome and faced with a set of alternatives, choose the one that yields the highest mathematical expectation of utility [Hershey and Baron, 1987]. The expected utility theory works in either direction: (1) Utilities can be inferred from choices in simple cases or assessed in a simple test and (2) these utilities can be used in the analysis of complex clinical cases. Utility assessment is based on two main principles:

1. Compensation A qualitative loss on one aspect of life can be compensated by a qualitative gain on another attribute.
2. Quantification Losses and gains can be expressed quantitatively.

Some fifteen years ago, utility assessment made its first cautious steps into medical decision making. Important work has been done by Torrance et al. in the development of simple utility assessment tests for clinical use, especially in the development of the basic reference gamble and the time tradeoff test [Torrance, 1987]. Practical and conceptual shortcomings of these methods actuate the search for alternative measurement techniques. In this chapter we will describe the principles of and some experiences with these two methods and introduce additive
**conjoint measurement** as an alternative method that can meet at least some of the shortcomings of the former methods.

### 5.2 The Life Standard Gamble (LSG)

#### 5.2.1. Principles of the Life Standard Gamble (LSG)

The 'basic reference gamble' or 'life standard gamble' (LSG) is directly based on the utility theory as originally described by Von Neuman and Morgenstern. Adapted to the medical field, the Von Neuman-Morgenstern principle states that the utility of a certain health-state is equal to the expected utility of a gamble between a worse and a better health-state.

A first application of the LSG in oncology has been described by McNeil et al. [1978]. A LSG was used to assess risk attitudes in 14 patients operated upon for lung cancer. They were retrospectively asked to choose between surgery and radiotherapy, considering that in the case of surgery they would have a better longevity than following radiotherapy, but with a higher short-term operative mortality.

Figure 5.1 schematically gives an example of the LSG in which a patient was asked to assess the amount of years, -the certainty equivalent-, that s/he considered equally desirable as a 50/50 gamble between a longevity of 25 years and immediate death. The mathematically expected number of life-years of the gamble equals 12.5 years.

![Figure 5.1 · Schematic representation of a life standard gamble for the choice between a 50/50 gamble of a longevity of 25 years (utility = 1.0) or immediate death (utility = 0.0) and the individually assessed certainty equivalent CE in years with a utility equal to the expected utility EU of the gamble (EU = 0.5) (see text).](image-url)
If the certainty equivalent is similar to the mathematical expected life-expectancy, a subject is said to have a *risk-neutral* attitude on life years. *Risk-aversity* on near life-years states that subjects are more reluctant to take a gamble with life when faced with a short life-expectancy than with life-years somewhere in a distant future. This interpretation was used by several authors to adjust utilities for the quality of life, assessed by a time tradeoff test (see §5.3). By varying the basic values in the gamble, a utility curve for the risk-attitude on life-years can be constructed such as illustrated by figure 5.2.

**Step I  Assessment of the certainty equivalent to a utility of 0.5**

By attaching a utility to the respective outcomes on a utility scale ranging from 0.0 for immediate death to 1.0 for a longevity of 25 years, the average or 'expected utility' (EU) of the gamble represented in figure 5.1 is 0.5. If a patient considers 12.5 certain life-years equal to the gamble, s/he is said to behave 'risk neutrally'. If s/he is already satisfied with less than 12.5 years, say seven certain life-years to avoid the gamble, s/he is said to behave 'risk-aversively'.

**Step II  Assessment of the certainty equivalent to a utility of 0.25**

The certainty equivalent to a utility of 0.25 is calculated by replacing the longevity of 25 years in the 50/50 gamble by the certainty equivalent to a utility of 0.5. For instance, the risk-averse patient mentioned above, is asked for the certainty equivalent to the 50/50 gamble between 'immediate death' and 'seven years'. If the certainty equivalent is less than the mathematically expected 3.5 years, say three years, the patient is still said to be 'risk-averse'.

![Figure 5.2](image-url) *Utility curves on life-years U(l)* The solid line gives the linear utility on the life-years curve for a subject with a risk-neutral attitude. The broken line gives a non-linear utility function on life-years denoting risk-aversity.
Step III: Assessment of the certainty equivalent to a utility of 0.75
The certainty equivalent to a utility of 0.75 is calculated by assessing the certainty equivalent to a 50/50 gamble between a longevity of 25 years (utility = 1.0) and the certainty equivalent to a utility of 0.5 (as assessed in step I). For instance: The risk-averse patient in steps I and II is asked for the certainty equivalent to the 50/50 gamble between ‘seven years’ and ‘twenty-five years’, being for instance fourteen certain life-years.

We did a pilot-study with 15 volunteers to reproduce the results of McNeil and to investigate the feasibility of the LSG for practical medical decision making.

5.2.2. Population and methods
Fifteen volunteers aged 18-64 (mean = 28 years), most of them academically educated, were personally interviewed and asked to imagine having lung cancer with a choice between surgery and radiotherapy, of which surgery offered a longer life-expectancy but a higher short term mortality. The test used is given in appendix C.1.

5.2.3. Results and discussion
Five out of fifteen were not able to finish the interview completely. Twelve out of fifteen were not able to approach life as a simple gamble or did not understand the rules of the game.

Almost all of them stated that they grasped the rules of the gamble but there was an equivocal reluctance to deal with life as a gamble and they therefore disapproved of the LSG. This gives rise to severe doubt about the validity of the LSG. Repeated disappointing results in a group of 12 general physicians and in a group of 20 medical students (not presented) and the negative results of other researchers forced us to renounce further research on the LSG [Llewellyn-Thomas et al., 1982; McNeil et al., 1982; Tversky and Kahneman, 1981].

5.3 The Time Tradeoff Test (TTT)

5.3.1. Principles of the Time Tradeoff Test (TTT)
The time tradeoff test (TTT) is used for the assessment of utilities for the quality of life by tradeoffs between quality of life and life-expectancy. One of the few applications in oncology described in the literature was presented by McNeil et al.
in 1981. This study deals with volunteers assessing the preference between radiotherapy and surgery in T3 laryngeal cancer, with radiotherapy offering better quality of speech but a shorter life-expectancy than surgery.

The time tradeoff test is based on the major assumption that the utility of a certain health-state A is proportional to the minimal amount of years a patient is willing to accept to live in the better health-state B compared with the (larger) number of life years in the worse state A.

For example: In chapter B.11, a subject is faced with the choice between radiotherapy and surgery for cancer of the vocal cords. Successful radiotherapy will lead to NATURAL speech and surgery implies an artificial speech with a longer life-expectancy than following radiotherapy. If a subject faced with a life-expectancy of 25 years with ARTIFICIAL speech is willing to tradeoff 5 years (i.e. leaving 20 years) to have a life with NATURAL speech, the utility of life with an ARTIFICIAL speech is said to be \( 1 - \frac{5}{25} = 0.8 \) of the utility of NATURAL speech.

The assumption of proportional tradeoffs requires that the utility of an ARTIFICIAL speech holds for all given life-years. For instance: If a subject faced with 25 years ARTIFICIAL speech is assumed to be prepared to sacrifice 5 years for a life with NATURAL speech, then this same subject is willing to sacrifice the same proportion (=20%) of 5 years ARTIFICIAL speech, i.e. 1 year. The present study was designed to test this condition of proportional tradeoffs in 65 medical professionals.

5.3.2. Population and methods
Fifty-six health-science students and nine general physicians were asked to answer a written interview. The interview began with a description of the goals and the hazards of both radiotherapy and surgery for cancer of the vocal cords. The subjects were asked to imagine they had cancer of the vocal cords and a choice between radiotherapy and surgery. It was suggested that the test was designed to help patients with cancer of the vocal cords to facilitate a treatment choice.

The time tradeoff test consisted of six items. In each item the subject had to assess the maximum number of life-years s/he was willing to sacrifice from a given life-expectancy with an artificial speech, in order to live with natural speech. The given life-expectancy in the six consecutive items was 25, 20, 15, 10, 5 and 2 years. It was explicitly stated that, if desired, it was possible to tradeoff parts of a year or to express the tradeoff in months.
The test was concluded with a short questionnaire on some personal characteristics (sex, age, marital state, education) and two multiple-choice questions investigating the ability to imagine having cancer and the ability to deal with quality of life in a quantitative way.

Linear regression statistics were used to test the condition of proportional tradeoffs: Letting \( L \) denote the length of life and \( U(a) \) the utility of artificial speech, the condition of proportional tradeoffs holds if in the regression equation \( U(a) = c + \beta L \) the coefficient \( \beta = 0 \) so that \( U(a) \) equals the constant \( c \).

### 5.3.3. Results

On a utility scale ranging from 0.0 for death to 1.0 for life with natural speech, figure 5.3 shows the utility of artificial speech as a function of life-expectancy. The broken horizontal line gives the expected graph for strict proportional tradeoffs for all life-expectancies. The solid line interpolates the mean utility, showing a decreasing utility attached to artificial speech for increasing life years (\( \beta = -0.003 \)). The graph shows a significant deviation from the condition of proportional tradeoffs (\( P < 0.01 \)). Neither sex, age, marital state, family circumstances nor smoking behaviour showed any particular association with these outcomes.

Almost all (95.4%) subjects stated that they had at least some difficulty to imagine they had cancer of the vocal cords and two-third (66.2%) had great problems or were unable to imagine having cancer of the vocal cords.

With respect to dealing with the quality of life quantitatively, 97% had at least some problems and 73.8% had severe problems or found it practically impossible.

Figure 5.3 · The utility of artificial speech \( U(a) \) according to the number of years to live as expected if the condition of proportional tradeoffs would hold (broken line) and as observed in 65 medical professionals (solid line) (see text).
5.3.4. Discussion
This study confirms the findings of a pilot study with volunteers (not presented) that the time tradeoff test reflects a subject’s actual attitude against quality of speech and survival; the non-patient population, although educated in clinical thinking and accustomed to medical problems-, encountered the greatest problems imagining having cancer of the vocal cords.

A major observation is that the TTT did not meet the required condition of proportional tradeoffs. This fact poses great questions to the usefulness in decision analysis: In individual cases, the eventual choice between radiotherapy and surgery may strongly depend on whether a subject is offered 25 years or 5 years for tradeoff and may less depend on the ‘real’ utility attached to the quality of speech.

Violation of the condition of proportional tradeoffs between the quality of life and life-years can be explained by two phenomena:

(1). **Mutual dependence of attributes.** The tradeoffs are proportional but the utility of artificial speech is determined by the utility of life-years, i.e. the shorter the length of life, the higher the utility of artificial speech. Without knowledge about the property, i.e. the mathematical relation-, of the dependence between quality of speech and quantity of life, it is not possible to correct for possible violations of proportionality. One can imagine some kind of dependence for extreme short lengths of life: When faced with only days or hours to live, one might at least wish to have a natural speech to be able to make final arrangements. In that case, the utility of artificial speech must be very low. In less extreme situations it is hard to explain disproportionality by mutual dependence of attributes. In general, independence between attributes is considered an axiom in utility assessment.

(2). **Non-constant utility function of life-years U(l),** i.e. not all years have an equal value. In general, life-years in the nearby future will be valued higher than distant life-years. Such a utility function much resembles the non-linear utility function on life-years due to risk-aversity $U(l)$ as depicted in figure 5.2 (dotted line). Then, the ordinate in figure 5.3 does not represent the utility of artificial speech, but a hybrid utility of two different (mutual independent) utility functions: One for the quality of speech $U(q)$ and one for life-years $U(l)$.

As discussed by Von Winterfeldt and Edwards [1986, p. 221 ff.], some utility theorists make a distinction between utility functions and value functions. Utility functions are assessed in gambles, such as the LSG, and value functions are
assessed in direct scaling methods in a 'riskless' context\(^2\), such as the TTT. We agree with Von Winterfeldt and Edwards that the distinction between utility functions and value functions is spurious and that these functions can be treated similarly in decision analysis. We further state that it is potentially harmful to correct quality of life utilities assessed in a TTT with utilities for life-years assessed in a LSG, as was suggested by McNeil et al. [1981] Although there is a semantic difference between trading and gambling with life-years [McNeil et al., 1978, Pliskin et al., 1980], we have doubts whether this difference holds in actual utility assessment.

A utility function of life-years with relatively higher utilities for life-years in the nearby future than for distant life-years, can be expressed by some kind of discount utility curve. For instance, a yearly discount of 10\% states that the utility of a certain year \(i\) is 10\% less than that of the preceding year \(i-1\). Such discount values were used by Glasziou and Simes [1988] in a quality-adjusted survival analysis in patients with breast cancer. Discount values for the utility of life-years are more based on assumptions than on empirical evidence.

5.4. Additive Conjoint Measurement (ACM)

5.4.1 Principles of ACM

The practical difficulties of subjects to deal with risks in the life standard gamble and the conceptual shortcomings of the time tradeoff test, require a search for alternative measurement techniques to assess separate utility functions for (a) the quality of life and (b) the value of life-years. Conjoint measurement models may meet the shortcomings of the previous methods by assessing these two utility functions simultaneously in a riskless context [Luce & Tukey, 1964]. Conjoint measurement models presume the existence of (a) separate utility functions for the quality of life \((U_Q)\) and life-years \((U_L)\) and (b) a conjunction of both attributes resulting in a conjoint utility function \(U_{Q,L}\) that can be described as a simple mathematical function based on the separate utility functions. The most simple mathematical model presumes a simple additive utility function

\[ U_{Q,L} = U_Q + U_L \quad \{5.3\} \]

\(^2\) Riskless in the sense that, -in the test-, there are no explicitly mentioned absolute risk expressions or probability statements. However, risk attitude or risk related behaviour may implicitly be considered in a subject's preference assessment.
Under certain conditions, the utility function $U_{Q,L}$ and the separate utility functions $U_Q$ and $U_L$ can be inferred from a preference ranking on pairs of life-expectancy $L$ and the quality of life $Q$. These pairs $Q,L$ are denoted as ‘dyads’. The model conditions are:

1. **Independence of attributes**: The utility of one attribute is not determined by the utility of the co-attribute. For instance, let $q, r, s$ denote the levels on attribute $Q$ and let $l, m, n$ denote the levels on attribute $L$. If you prefer the dyad $(q,l)$ to $(r,l)$, you prefer $q$ to $r$ for any other level on $L$. Similarly, if you prefer $(q,l)$ to $(q,m)$ then you prefer $l$ to $m$ for any level on $Q$.

2. **Transitivity of preferences**: If for any dyads $A,B$ and $C$, you prefer dyad $A$ to $B$, and $B$ to $C$, then you prefer $A$ to $C$. The opposite (or violation) of transitivity is called circularity. For instance, if you prefer dyad $A$ to $B$, $B$ to $C$ and $C$ to $A$, there is a cycle in the preference ranking on the three elements $A,B$ and $C$. In a preference ranking on the dyads $A,B,C .$, both the number of cycles and the number of dyads in each cycle contribute to the severity of intransitivity [Bezem, 1981].

3. **Double cancellation of preferences**:

<table>
<thead>
<tr>
<th>L</th>
<th>Q</th>
<th>and if:</th>
<th>then:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>mute</td>
<td>m,3</td>
<td>m,6</td>
<td>m,9</td>
</tr>
<tr>
<td>oesoph.</td>
<td>o,3</td>
<td>o,6</td>
<td>o,9</td>
</tr>
<tr>
<td>normal</td>
<td>n,3</td>
<td>n,6</td>
<td>n,9</td>
</tr>
</tbody>
</table>

Figure 5.4 • *Double cancellation demonstrated by a 3x3 matrix for the dyads of $Q$ (mute, oesophageal speech and normal speech) and $L$ (3, 6 and 9 years) (see text)*

Let **mute**, **oesophageal** and **normal** denote three levels on quality of speech $Q$, and let 3, 6 and 9 years denote three levels on life-expectancy $L$. Figure 5.4
gives the dyads of \((Q,L)\) ordered in a 3x3 matrix. If you prefer the dyad \((n,6)\) to \((o,3)\) (left panel) and the dyad \((o,9)\) to \((m,6)\) (middle panel) then you prefer the dyad \((n,9)\) to \((m,3)\) (right panel). If the condition of double cancellation is violated, there is no solution that yields additive utility functions from the preference ranking. With \(q \geq 3\) and \(l \geq 3\) the number of levels in \(Q\) and \(L\), respectively, the condition of double cancellation goes for any 3x3 submatrix resulting from the original \(q \times l\) matrix by picking three levels on \(Q\) and three levels on \(L\).

### 5.4.2 Determining a Preference Ranking

Based on ACM, a test was designed to assess an individual preference ranking on a set of dyads of quality of speech and life-expectancy in patients with cancer of the vocal cords.

In our experiments, we chose five \((n\) and \(m)\) levels on each attribute. On life-expectancy, 3, 6, 9, 12, and 15 years. On quality of speech, Mute (M), electronic larynx (E), oesophageal speech (O), hoarseness (H), and normal speech (N). The five levels for the quality of speech might need some clarification. By surgery for cancer of the vocal cord the voice box of a patient is sacrificed. These patients have to learn an artificial speech. About 70% of these patients manage to learn a so-called oesophageal speech by belching swallowed air. This method requires patience, intelligence, and long practice and not all patients are successful. The latter patients (25%) may be helped by a so-called electronic larynx. This electronical device produces a vibration that is amplified and modelled by the oral cavity. The result is a somewhat metallic and monotonous speech. About 5% of patients will never learn any form of artificial speech and stay mute for the rest of their life. Following successful radiotherapy, the quality of speech may range from near normal speech to hoarseness. From five life-expectancies and five qualities, 25 dyads can be formed ranging from worst, being 3 years mute, to best, being 15 years normal speech. The ranking of the other 23 dyads has to be assessed by individual testing. Figure 5.5 gives an example of a preference ranking of the 25 dyads in a 5x5 matrix (\(n \times m\) data matrix).

Each cell contains the rank number for each dyad of \(Q\) and \(L\). Having these 25 dyads directly put into a preference ranking results into a transitive ranking but this task mostly appears very difficult if not virtually impossible for most subjects. Also, the result will generally not be very stable. It is much easier for a subject to
make a choice between two simple alternatives. Therefore we made pairs of dyads, -so-called paired comparisons-, and we asked the subject to assess his or her preference in each paired comparison.

<table>
<thead>
<tr>
<th>Q</th>
<th>L.</th>
<th>3y</th>
<th>6y</th>
<th>9y</th>
<th>12y</th>
<th>15y</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mute</td>
<td>0</td>
<td>4</td>
<td>8</td>
<td>13</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>Electronic larynx</td>
<td>1</td>
<td>6</td>
<td>11</td>
<td>16</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>Oesophageal</td>
<td>2</td>
<td>7</td>
<td>12</td>
<td>17</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>Hoarseness</td>
<td>3</td>
<td>9</td>
<td>14</td>
<td>19</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>5</td>
<td>10</td>
<td>18</td>
<td>22</td>
<td>24</td>
<td></td>
</tr>
</tbody>
</table>

Figure 5.5 - Arrangement of a preference ranking on 25 dyads of quality of speech and life-expectancy in a 5x5 matrix (n x m data matrix).

For example:

What do you prefer?
A. Living 9 years with a normal speech
   or
B. Living 12 years mute

From 25 dyads, $25^2 = 625$ paired comparisons can be constructed. Figure 5.6 gives all 625 paired comparisons arranged in a 25x25 matrix (nm x nm matrix). The entry in each matrix cell is 0 or 1, indicating that either the column dyad (1) or the row dyad (0) is preferred. Under certain constraints, preferences need not be assessed on all 625 paired comparisons:

1. If we presume the following natural preference order on the attributes, i.e
   for life-expectancy: $15 > 12 > 9 > 6 > 3$ years
   and for the quality of speech: $N > H > O > E > M$,
then only those comparisons are non-trivial in which the dyads show opposite directions of this natural order of the attributes. The \( > \) sign denotes strict preference in the observed data.

Figure 5.6 • Arrangement of 625 paired comparisons in a 25x25 matrix. This matrix gives an empirical example: One ('1') denotes preference of a row-dyad to the corresponding column-dyad. \( \Sigma c \) denotes the column sum-score, \( \Sigma r \) denotes the row sum-score. Non-trivial paired comparisons have been represented in bold characters (see text). (To enhance readability, zero's in trivial paired comparisons have been replaced by '-' .)
For instance in the paired comparison:

A. Living 9 years with a normal speech  
   or  
B. Living 12 years mute

the order of attributes is opposite, since 9 years < 12 years and normal speech > mute. In each non-trivial paired comparison a patient thus chooses either the better quality of speech or the better life-expectancy. In the 625 paired comparisons, 200 are non-trivial.

<table>
<thead>
<tr>
<th></th>
<th>3</th>
<th>3</th>
<th>3</th>
<th>3</th>
<th>3</th>
<th>6</th>
<th>6</th>
<th>6</th>
<th>6</th>
<th>6</th>
<th>9</th>
<th>9</th>
<th>12</th>
<th>Σr</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>M</td>
<td>E</td>
<td>O</td>
<td>H</td>
<td>N</td>
<td>M</td>
<td>E</td>
<td>O</td>
<td>H</td>
<td>M</td>
<td>E</td>
<td>M</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>3M</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0...</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>3E</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0...</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>3O</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0...</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>3H</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0...</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>3N</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0...</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>6M</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0...</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>6E</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0...</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>6O</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0...</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>6H</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0...</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>9M</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0...</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td></td>
<td>9E</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0...</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td></td>
<td>12M</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1...</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td></td>
<td>9O</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1...</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sc</td>
<td>24</td>
<td>23</td>
<td>22</td>
<td>21</td>
<td>19</td>
<td>19</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>16</td>
<td>15</td>
<td>14</td>
<td>13</td>
</tr>
</tbody>
</table>

Figure 5.7 · *Detail from the binary matrix in figure 5.6 sorted by increasing row sum-scores and by decreasing column sum-scores. The paired comparisons in italics correspond with the dyads involved in a cycle (see text).*
2 The 200 non-trivial paired comparisons may easily be reduced to 100. For instance the paired comparison

A Living 9 years with a normal speech
  or
B Living 12 years mute

is equivalent to the paired comparison

A Living 12 years mute
  or
B Living 9 years with a normal speech

So, 100 paired comparisons suffice for representing all non-trivial (hard) choices. Appendix C.3 gives the 100 paired comparisons used in the ACM test. On each paired comparison, a subject was asked to indicate a preference either for a better quality of speech (dyad A) or for a longer life (dyad B).

Appendix C.4 gives the FORTRAN programme to generate all paired comparisons from a matrix with \( n \) levels on \( Q \) and \( m \) levels on \( L \) and to check the \( nm \times nm \) matrix for (violation of) transitivity. Figure 5.7 gives a detail from the binary matrix of figure 5.6 sorted by increasing row sum-scores and by decreasing column sum-scores. The sorted matrix graphically displays an asymmetric structure around the blank diagonal line. The asymmetry is due to circularity in four dyads: 3N, 6M, 6E, 6O.

Based on this binary matrix, Bezembinder designed a mathematical method to identify the cycles in a preference ranking [Bezembinder, 1981]. It can be shown that in a transitive binary \( nm \times nm \) matrix, sorted by increasing row sum-score, the cumulative row sum-score \( S_{(k)} \) for dyad \( k \) equals \( k(k-1)/2 \) for \( k = \{1, 2, \ldots, nm\} \).

Further, it can be proven that a dyad for which \( S_{(k)} > k(k-1)/2 \) is involved in a cycle. In a preference ranking on dyads, all consecutive dyads with \( S_{(k)} > k(k-1)/2 \) and the next dyad with \( S_{(k)} = k(k-1)/2 \) belong to one cycle. For instance, the next preference ranking on dyads shows that dyads 3N, 6M, 6E and 6O all belong to one and the same cycle.

---

3 For a matrix with \( n \) levels on \( Q \) and \( m \) levels on \( L \), the number of dyads is \( \frac{n \times m(n-1)(m-1)}{4} \).
In a single test, circularity may occur accidentally because a subject may simply have made a mistake by putting one of the test-cards on the wrong stack. Therefore, the ACM-test is performed three times with one week intervals. If for a specific dilemma Q is preferred at least twice, Q is said to be preferred for this dilemma. Otherwise it is said that L is preferred. Then, the results of all three replications can be processed as if it were one single test.

Circularity after three replications of the test is likely to be explained by violation of transitivity in the preference ranking. Both the number of cycles in a preference ranking and the number of dyads in each cycle contribute to the severity of violation of transitivity in a preference ranking [Bezembinder, 1981].

A simple method to restore transitivity is by switching or 'reversing appropriate choices, the well-known solution is the one that can be obtained by the smallest number of switches. However, this method does not yield a unique solution [Bezembinder, 1981].

Since we want to obtain a patient’s 'real' preferences, we suggest the practical solution to confront a subject with his/her circularity and have him/her correct his/her preferences by ordering the dyads involved in a cycle. For instance, Figure 5.7 shows a cycle for the dyads 3N, 6M, 6E and 60. In a discussion between the interviewer and the subject, the subject is asked to rearrange the four dyads in a transitive order.

For practical reasons, these corrections are only made for limited violations of transitivity, involving only few and small cycles. Of course, this procedure may also lead to the conclusion that the patient is unable to reach a transitive order. In that case, this procedure does not result into an advice as to what treatment is to be preferred.
5.4.3 Obtaining an Additive Preference Ranking

In the former paragraph, we described how to obtain a transitive preference ranking on dyads of Q and L. Next we have to determine whether there is an additive solution for this preference ranking, i.e., whether we can infer two independent utility functions \( U_Q \) and \( U_L \) as well as the additive conjoint utility function \( U_{Q,L} = U_Q + U_L \). Further, if the preference ranking meets the conditions of 'independence of attributes' and 'double cancellation of preferences', there generally will be an additive solution.

Since in the ACM-test strict order was presumed on both attributes Q and L, independence requires that an nxm data matrix, such as figure 5.5, shows increasing cell entries in each column 1,2 \( n \) for each row, and increasing cell entries in each row 1,2 \( m \) for each column. In this case, a check on independence can readily be seen by a visual inspection of the data matrix.

Figure 5.4 shows the check on double cancellations in a 3x3 data matrix. In an nxm data matrix, each 3x3 submatrix should be tested for double cancellation in each of its \((3!)^2 = 36\) possible permutations. Any nxm matrix has \( (n!/3(n-3)!)(m!/3(m-3)!)) \) possible 3x3 submatrices, a 5x5 matrix has 100 possible 3x3 submatrices. Double cancellation has to be checked for any permutation of rows and columns of the 3x3 matrix, yielding \((3!)^2 = 36\) permutations. Then, a 5x5 matrix requires 3600 checks on double cancellation if each of \( n \) rows and each of \( m \) columns are fixed, such as in figure 5.4 and figure 5.5. 36 checks are required for each 3x3 submatrix.

Analogously to the correction of the violation of transitivity, one can suggest to correct violations of independence and of double cancellation in a discussion with the subject. However, the limited experience with the ACM-test almost always showed at least some intransitivities. Correcting these intransitivities always turned out to be sufficient for establishing both independence and double cancellation.

5.4.4 Inferring Utility Function \( U_{Q,L} \), \( U_Q \) and \( U_L \)

Two methods are available to infer utility functions \( U_{Q,L} \), \( U_Q \) and \( U_L \).

The most straightforward method is to find a solution by trial and error. Such an iterative process is performed by the FORTRAN-program UNICON [Bendermacher & Thissen, 1987]. UNICON is based on an algorithm which looks for a solution by minimizing the error or 'stress function'. For an additive preference ranking on dyads, a solution can be found without any stress. In that case, namely...
when additivity is completely met, there generally is more than one unique solution. Following an iterative search process, UNICON gives only one, quite arbitrary solution, namely the first solution with stress = 0.

Since there are many solutions to an additive preference ranking, it is essential to define the boundaries (or extreme solutions) of the solution space and to find a representative solution within this space. ORDMET3 is a FORTRAN program by Roskam which finds the extreme solutions by describing and solving the preference ranking as a set of linear inequalities [Roskam, 1987]. A set of linear inequalities is a consistent set of linear comparisons (≤ or ≥) with unknown elements which can be solved by linear programming presuming an additive model. For instance: The 3x3 matrix in figure 5.8 yields 36 linear inequalities:

\[
\begin{array}{ccc}
  a & b & c \\
  q & 1 & 2 & 4 \\
  r & 3 & 5 & 7 \\
  s & 6 & 8 & 9 \\
\end{array}
\]

<table>
<thead>
<tr>
<th></th>
<th>q+a &lt; q+b</th>
<th>q+a &lt; q+c</th>
<th>q+a &lt; r+a</th>
<th>q+a &lt; r+b</th>
<th>q+a &lt; r+c</th>
<th>q+a &lt; s+a</th>
</tr>
</thead>
<tbody>
<tr>
<td>q+b &lt; q+c</td>
<td>q+b &lt; r+a</td>
<td>q+b &lt; r+b</td>
<td>q+b &lt; r+c</td>
<td>q+b &lt; s+a</td>
<td>q+b &lt; s+b</td>
<td></td>
</tr>
<tr>
<td>q+c &lt; r+a</td>
<td>q+c &lt; r+b</td>
<td>q+c &lt; r+c</td>
<td>q+c &lt; s+a</td>
<td>q+c &lt; s+b</td>
<td>q+c &lt; s+c</td>
<td></td>
</tr>
<tr>
<td>r+a &lt; r+b</td>
<td>r+a &lt; r+c</td>
<td>r+a &lt; s+a</td>
<td>r+a &lt; s+b</td>
<td>r+a &lt; s+c</td>
<td></td>
<td></td>
</tr>
<tr>
<td>r+b &lt; r+c</td>
<td>r+b &lt; s+a</td>
<td>r+b &lt; s+b</td>
<td>r+b &lt; s+c</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r+c &lt; s+a</td>
<td>r+c &lt; s+b</td>
<td>r+c &lt; s+c</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>s+a &lt; s+b</td>
<td>s+a &lt; s+c</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>s+b &lt; s+c</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

If the 3x3 matrix meets additivity (transitivity, independence and double cancellation), the eight linear inequalities in the first column are sufficient to determine the solution space, i.e. to determine all sets of utility values q, r, s, a, b, and c. The eight linear inequalities are the comparisons between each cell-entry and its immediate successor in the strict simple order given by the 3x3 matrix. By fixing one utility value on either attribute, -generally the worst Q and the least L-, the solution space can be described as a convex polyhedral cone within a space.
with \((n-1)(m-1)\) dimensions or degrees of freedom. The extreme solutions can then be imagined as edge-vectors of the convex polyhedron space (see figure 5.9). ORDMET3 allows one to follow one's own predilection and to choose any solution within this space as a representative solution. Three methods, each yielding only one 'unique' solution within the solution space are mentioned here:

1. The maximin solution, representing the vector which correlates best with any of the extreme solutions (the smallest angle with any of the extreme vectors) [Roskam, 1987];
2. The least squares solution representing the solution with the smallest distance to any extreme solution [Tversky & Zivian, 1966].
3. The centroid solution representing the average of the extreme solutions. Chapter B.12 describes how the centroid utility function \(U_Q\) and \(U_L\) can be used in a decision tree analysis.

Figure 5.9 • A convex polyhedral cone bound by four edge vectors.

5.4.5. Limitations of ACM; Further research

Table 5.1 summarizes some characteristics of additive conjoint measurement compared with the life standard gamble and the time tradeoff test.

Independence of attributes is prerequisite for all measurement models here considered. Transitivity is presumed both for the TTT and the LSG without considering whether or not transitivity is met; the ACM-model provides both a tool to check for transitivity, to correct for (limited) violations of transitivity and to infer
independent utility functions for the quality of life and for life-expectancy from a transitive preference ranking on dyads of quality and quantity of life.

Proportionality of tradeoffs between the quality and quantity of life are both required in the TTT test and in the LSG, but not in the ACM test. In the LSG test, the tradeoffs and the resulting utilities are further disturbed by a subject's attitude towards risks and probabilities, hence, the LSG is generally not used for the assessment of tradeoffs between the quality and quantity of life but merely for the assessment of risk attitude upon life-years [see Pliskin et al., 1980, McNeil et al., 1978, 1981]. Neither in the TTT test nor in the ACM test are risk properties explicitly mentioned in the assessment of tradeoffs. However, such risk properties may implicitly be considered by a subject during his/her utility assessment. Adjustment of utilities for the quality of life, as assessed in a TTT test by risk attitude upon life-years assessed in a LSG, as proposed by Pliskin et al. and by McNeil et al., may falsely lead to overadjusted utilities for the quality of life.

Table 5.1: Summary of some characteristics of additive conjoint measurement (ACM), the life standard gamble (LSG) and the time tradeoff test (TTT)

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>ACM</th>
<th>LSG</th>
<th>TTT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independence presumed</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Transitivity</td>
<td>Required</td>
<td>Presumed</td>
<td>Presumed</td>
</tr>
<tr>
<td>Proportional tradeoffs</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Risk attitude</td>
<td>Neutral</td>
<td>Averse or</td>
<td>Neutral</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Seeking</td>
<td></td>
</tr>
<tr>
<td>Practical ease</td>
<td>Good</td>
<td>Poor</td>
<td>Poor</td>
</tr>
<tr>
<td>Mathematical ease</td>
<td>Complex</td>
<td>Bedside</td>
<td>Bedside</td>
</tr>
</tbody>
</table>

A major difficulty of the LSG, and to a minor extent of the TTT test, is that subjects encounter great difficulties to grasp the quintessence of the gamble or to deal with the quality of life in a quantitative way. The simplicity of simple two-by-two choices...
in the ACM test may facilitate to establish quantitative tradeoffs of subjective values of quality of life and life-expectancy.

The calculation of utilities are quite simple for both the TTT test and the LSG, these methods have therefore been promoted as simple bedside methods to assess utilities. The ACM test requires more effort: The test itself takes only 15 to 20 minutes, but has to be replicated three times and generally requires a fourth time to correct for violations of the model assumptions. Both the identification of violations of the model assumptions, the inference of additive utility functions $U_Q$ and $U_L$ and the decision tree analysis using time dependent utilities, have been facilitated by specific computer programmes.

5.6 References


18 Roskam EE. ORDMET3: An improved algorithm to find the maximin solution to a system of linear (in)equalities. Internal report 87 MA 06, Nijmegen Institute for Cognition-research and Information-technology (NICI), Nijmegen, 1987.


24 Verkes RJ. *De rationaliteit van beslissingsanalyse bij medisch therapeutische beslissingen, in het bijzonder in de oncologie*. Doctoral thesis. Vakgroep Metamedica, Faculty of Medicine, State University Leyden, 1987.


clinical decision making in oncology

part b:
applications
6. Results of radiotherapy and surgery for glottic carcinoma

6.1 Introduction

Cancer of the larynx accounts for 2% of malignancies diagnosed in the United States. It is a predominantly male disease. The estimated incidence for 1983 was 9,100 in men and 1,800 in women with an estimated death rate of 3,100 in men and 600 in women [5,50].

About 60% of the patients are older than 50 with a median age of 60 [52]. Smoking is strongly associated with laryngeal cancer, with increased risks up to 25 times found for war-veterans smoking more than 40 cigarettes daily compared with non-smokers [27,5].

Laryngeal cancer originates in the glottic (55%), supraglottic (40%) and subglottic regions (5%). Patients with glottic carcinomas are curable both by radiotherapy and by surgery. The good prognosis can be attributed to:

1. Early detection of the tumour because of early symptoms, especially hoarseness,
2. Late metastatic tumour spread because of the sparse lymphatic and vascular supply of the vocal cords,
3. Relatively easy technical approach of the tumour by surgery or radiotherapy.

Although the survival rates are favourable, the disease itself as well as its treatment can severely affect glottic anatomy and functioning, resulting in impaired voice qualities ranging from simple hoarseness to total loss of speech. Surgery, even so-called conservation surgery, gives more voice-debilitating results than radiotherapy [33,7,20].

Currently, radiotherapy is accepted as the first mode of treatment for glottic carcinoma confined to the vocal cord (T₁N₀M₀). Most therapeutic centres favour radiotherapy in the case of tumour extension to either the supraglottic or subglottic regions without fixation of the vocal cords (T₂N₀M₀) [10,58,38,37,8,34,49,11,7,43,15,23,39]. Others advocate surgery as first treatment, especially if the mobility of the vocal cords is impaired [3,4,29,32,35,44]. Radiotherapy may also be the primary mode of treatment for patients with more

---

1 L J A Stalpers, MD, A L M Verbeek, MD, PhD and W A J Van Daal, MD, PhD. Published in Cancer Treatment Reviews, 1987, 14:131-141.
extensive tumours ($T_3$ and $T_4$), especially since the understanding of radiotherapeutic failures has improved and irradiation techniques have been optimized [9, 14, 16, 18, 19, 43, 45, 56].

In a 1970 literature review on the role of radiotherapy for cancer of the larynx Vermund stated: 'It is only by constant comparison and review of the various modes of treatment that we can hope to achieve better results'. Especially when treatment modalities are controversial, treatment results as presented in scientific papers, have to be carefully reviewed and compared.

The aim of this study was to see whether, by reviewing the literature, we could find factors and arguments that would favour radiotherapy or surgery for glottic cancer. This is essential when more extensive tumours are concerned for which the optimal choice of treatment is still controversial.

6.2 Data and methods

The laryngologic and oncologic literature of the period 1973-1986 has been reviewed for factors and arguments which play a decisive role in the choice between primary radiotherapy and primary surgery, as well as for treatment results of glottic carcinoma.

Only articles using the TNM system for tumour classification were studied [1, 25]. Because the degree of mobility in $T_2$ carcinomas is assumed to have prognostic significance, a distinction was made between $T_{2a}$ and $T_{2b}$. This literature study only deals with treatment results of tumours that had not metastasized to regional lymph nodes or distant organs at the time of primary diagnosis.

We described treatment results in terms of 5-year survival rates, 5-year disease-free survival rates, recurrence rates and salvage rates. The 5-year period was chosen because it is used in most articles and because the few authors describing longer periods of investigation, state that their survival results do not essentially differ from 5-year results [22, 43, 51, 53]. Shorter periods (e.g. 3 years) were not evaluated for survival and disease-free survival because data were scarce. In all, 26 articles were reviewed describing 5-year survival and 5-year disease-free survival; for the study of recurrence and salvage after recurrence 28 articles were evaluated.
According to the UICC TNM general rules, survival results can be presented in two ways [24]:

1. The \textit{crude survival rate} calculated by the direct method, is the number of persons known to be alive at the end of the period considered, divided by the total number who were alive at the beginning of this period:

\[
S_{\text{crude}} = \frac{\text{alive}}{\text{alive} + \text{dead} + \text{lost}}
\]

The alternative or life table method of computing survival rates uses all the information accumulated over the whole period of observation and provides a description of the survival pattern [2,6,28]. In describing five-year survival, an actuarial method gives a weighted correction for patients who left the study alive at the end of the study-period and had no five-year follow-up. For these patients the term withdrawal is used.

2. The \textit{corrected or adjusted survival rate}. There are different methods of correcting the crude rate for death from intercurrent diseases and for those patients lost to follow-up. In an actuarial method these types of corrections can be made by regarding those patients as withdrawals. This methodic analogy of correction may lead to the erroneous assumption that an actuarial method always gives a correction for death from intercurrent diseases and for patients lost to follow up. This explains why many authors seem to compare an 'actuarial' method with a 'crude' method, while in fact they are comparing a 'corrected actuarial' method with an 'uncorrected actuarial' method in the sense described above. In this study 'corrected' rates are rates from studies where correction was made for death from intercurrent disease and for patients lost to follow-up. If it was not clear from the study whether any or what correction was made, a 'crude' rate was assumed.

Salvage rates and disease-free survival rates are given in ranges and means. These means are proportionally weighted for the number of patients in every study:

\[
\text{mean} = \frac{\sum (n_i \times p_i)}{\sum n_i}, \quad n_i = \text{number of eligible patients in study } i, \quad p_i = \text{survival rate in study } i
\]
In describing recurrence rates and salvage rates, cumulative means are given:

\[
\text{cumulative mean} = \frac{\sum_i (r_i)}{\sum_i n_i},
\]

where \( r_i \) is the number of recurred (or salvaged) patients in study \( i \) and \( n_i \) is the number of patients at risk in study \( i \).

Salvage was defined as curation after recurrence, generally by surgery. Using this definition, two kinds of salvage rates are described in the literature. The first salvage rate is the rate of successful salvages among all patients who had tumour recurrence. The second is the rate of successful salvages among all patients who had undergone a salvage procedure. The second rate disregards patients who had tumour recurrence but, for whatever reason, did not undergo a salvage procedure. If possible, salvage rates of both types will be presented.

### 6.3 Results

Radiotherapeutic schedules and techniques differ within time and within the varying treatment centres. Table 1 summarizes irradiation dose, overall treatment time, dose-fractionating schedules, techniques, and some indicative survival figures as described by the different authors. Articles in which no technical specification for radiotherapy were mentioned are not tabulated in Table 1 [32,37,47].

**Survival and disease-free survival (Tables 2 and 3)**

5-year survival results for non-metastasized glottic carcinoma are summarized in Table 2. For \( T_1N_0M_0 \) tumours, crude and corrected data are quite different. A similar difference between crude and corrected figures can be found in the review of \( T_2N_0M_0 \) carcinomas, although surgical results are scarce, as are survival results for \( T_3 \) and \( T_4 \) tumours. The broad ranges in survival for \( T_4 \) tumours can be ascribed to the small populations, both for radiotherapy and for surgery. Table 3 summarizes data on 5-year disease-free survival. Unlike radiotherapeutic results, only few surgical results have been published.
<table>
<thead>
<tr>
<th>Reference + Author (year of publication)</th>
<th>Class.</th>
<th>Dose (Gy)</th>
<th>Time (weeks)</th>
<th>Fractionation (no/week,dose)</th>
<th>Technique + field size</th>
<th>5-year Survival</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Crude</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>T1</td>
</tr>
<tr>
<td>4 Brandenburg (1977)</td>
<td>T1-4</td>
<td>60-70</td>
<td>6-7</td>
<td>-</td>
<td>Co; &gt;T1 cervical nodes=50 Gy</td>
<td>-</td>
</tr>
</tbody>
</table>
| 7 Dickens (1983)                        | T1-2   | 56-60     | 5-7         | -; 1.73-2.50 Gy               | Co; lateral fields + wedge | 82% | 63% | 54% | 75% | -
| 8 Ennuyer (1975)                        | T1-4   | 60-75     | 4-6.5       | -                             | Co; individual dose by response | 90% | 80% | -   | -   |
| 11 Goffinet (1973)                      | T1-4   | 50-70     | 5-7         | 4-5/week; 2.0 Gy              | 4MeV; lateral wedged fields 4x4 or 4x5 | -   | 36% | -   | -   |
| 12,15 Harwood (1979)                    | T1-2   | 55        | 5           | 24-26 fx                      | Co; compression technique 5x5 or 6x6 | 74% | -   | -   | -   |
| 16,18 Harwood (1980, 1981)              | T3-4   | 55        | 4-5         | 5/week; 25 Gy                 | Co; angled down + wedges + cast | -   | -   | 74% | 64% |
| 22 Hintz (1983)                         | T1     | -         | -           | -                             | Co; 4MeV, lateral angled wedged fields | 92% | -   | -   | -   |
| 23,49 Hunter (1980); Stewart (1975)     | T1-4   | ca 55     | 3           | 16 fx                         | 4MeV, antero-lateral wedged or parallel opposed fields | 73% | 59% | 47% | 33% |
| 29,30 Kaplan (1983, 1984)               | T1-4   | 60-65     | 6-6.5       | -                             | -; parallel opposed fields 5x5 to 6x6 or anterior oblique wedged fields | -   | -   | -   | 88% |
| 31 Kim (1978)                           | T1-2   | 50-70     | -           | 5/week; 2.25 or 2.5 Gy        | Co; parallel opposed fields 6x6 | 96% | -   | -   | -   |
| 35 Lippi (1984)                         | T1-2   | >50       | -           | 5/week; 1.8-2.0 Gy            | MeV, Co, 1 anterior electron or 2 opposed Co fields | 75% | 46% | -   | -   |
| 37,55 Miller (1975); Wang (1974)        | T1-4   | 65-70     | 6-7         | 5/week; 2.0 Gy                | 2MeV, Co, single or opposed lateral or wedged fields | -   | -   | 87% | 72% |
| 39 Mills (1979)                         | T1-2   | 55        | 5.5         | 5/week; 2.0 Gy                | Co; parallel opposed + wedged fields | 92% | 62% | -   | -   |
| 40 Minja (1984)                         | T1     | 30-65     | -           | 5/week, 1.4-2.0 Gy            | 250kV Orthovolt             | -   | -   | -   | -   |
| 41 Nass (1976)                          | T1-4   | 60-65     | 6-7         | 5/week; 2.0 Gy                | Co; parallel opposed fields 6x8 | -   | -   | -   | 82% (N.E.D.) |
| 43 Notter (1984)                         | T1-4   | 60-70     | 6-6         | 5/week; 2.0 Gy                | 2MeV, Co, 5MeV; opposed wedged fields | 85% | 76% | -   | -   |
| 51 Sung (1979)                          | T1-2   | 47-76     | 5-8         | 5/week; 2.0 Gy                | Orthovolt, Co, 6MeV, 22MeV; parallel opposed fields: 5x6 | -   | -   | 76% | 66% (N.E.D.) |
| 53 Van den Bogaert (1980)               | T2     | 40-70     | 4-7         | 5/week; 2.0 Gy                | Co; parallel opposed fields | 48% | -   | -   | -   |

fx = fraction(s), Co = Cobalt gamma; MeV = megavolt photons, Orthovolt = low energy photons, N E D = no evidence of disease
<table>
<thead>
<tr>
<th>Radiotherapy</th>
<th>Surgery</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3 • *Five-year no-evidence of disease (N E D.) after radiotherapy or surgery for non-metastasized glottic carcinoma*

<table>
<thead>
<tr>
<th></th>
<th>Radiotherapy</th>
<th>Surgery</th>
<th>References (patients at risk)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>crude</td>
<td>corrected</td>
<td>crude</td>
</tr>
<tr>
<td>T1</td>
<td>72% and 73%</td>
<td>76% - 93%</td>
<td>95%</td>
</tr>
<tr>
<td></td>
<td>mean = 72%</td>
<td>mean = 84%</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T2</td>
<td>37% - 72%</td>
<td>67% - 75%</td>
<td>59% and 61%</td>
</tr>
<tr>
<td></td>
<td>mean = 63%</td>
<td>mean = 70%</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T2a</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
</tr>
<tr>
<td>T2b</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
</tr>
<tr>
<td>T3</td>
<td>26% and 38%</td>
<td>51%</td>
<td>79%</td>
</tr>
<tr>
<td></td>
<td>mean = 30%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T4</td>
<td>25% and 33%</td>
<td>56%</td>
<td>50% and 67%</td>
</tr>
<tr>
<td></td>
<td>mean = 27%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

NP = not published

Hence, no differentiation between crude or corrected rates could be made. This may explain the broad range of recurrence rates. A cumulative mean rate of all articles reviewed has been calculated. For all stages recurrence after radiotherapy proved higher than after surgery. As 5-year corrected survival for T1 and T2 tumours is almost equal for either treatment, the discrepancy between recurrence rates and survival rates must to some extent be counterbalanced by better salvage results for radiotherapeutic failures than for surgical failures. This can be seen in Table 5. Salvage rates of the first type are lower than those of the second type. The discrepancy can run up from 3% for recurrences of T1 tumours to 21% for T3 tumours.
### Table 4  
**Recurrence rates for primary treated non-metastasized glottic carcinoma (not corrected)**

<table>
<thead>
<tr>
<th>Stage</th>
<th>Radiotherapy</th>
<th>Surgery</th>
<th>References (patients at risk)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>range</td>
<td>cumulative mean</td>
<td>range</td>
</tr>
<tr>
<td>T1</td>
<td>5% - 46%</td>
<td>335/2438 = 14%</td>
<td>0% - 18%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SU 3(32),29(31),31(17),35(22),</td>
</tr>
<tr>
<td>T2</td>
<td>16% - 60%</td>
<td>105/353 = 29%</td>
<td>11% - 31%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SU 3(58),4(9),8(18),29(24),35(62),47(63)</td>
</tr>
<tr>
<td>T2a</td>
<td>8% - 67%</td>
<td>17/82 = 20%</td>
<td>N P</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SU 29(9)</td>
</tr>
<tr>
<td>T2b</td>
<td>27% - 50%</td>
<td>19/61 = 31%</td>
<td>N P</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SU 29(24)</td>
</tr>
<tr>
<td>T3</td>
<td>35% - 83%</td>
<td>108/260 = 42%</td>
<td>0% and 31%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SU 4(49),8(8)</td>
</tr>
<tr>
<td>T4</td>
<td>66% - 75%</td>
<td>N P</td>
<td>31% - 50%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SU 4(13),8(9),26(32)</td>
</tr>
</tbody>
</table>

N P = not published

### 6.4 Discussion

Comparing treatment results of competing therapeutic modalities is necessary but difficult. The great variety of statistical methods that have been used in describing 5-year survival of primary treated glottic carcinoma makes comparison of radiotherapy and surgery extremely difficult. By making a division between crude and corrected survival rates, we could partly overcome such problems. As shown above, corrected rates are considerably more favourable than crude ones, which is not surprising since the disease manifests itself in older age when death from
cardiovascular and other diseases is substantial. Survival rates show almost equal results for primary radiotherapy of $T_1$ and $T_2$ tumours and primary surgery. The scarce data for $T_3$ and $T_4$ tumours suggest the same. Then, in $T_2$, $T_3$ and $T_4$ tumours opting for radiotherapy in preference to surgery may depend on the value that a patient attaches to the quality of voice [20,36,48].

### Table 5 · Cumulative mean rates for salvage of recurrence after primary radiotherapy or surgery of non-metastasized glottic carcinoma

<table>
<thead>
<tr>
<th></th>
<th>Radiotherapy salvage rate</th>
<th>Surgery salvage rate</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1st type</td>
<td>2nd type</td>
<td>1st type</td>
</tr>
<tr>
<td>$T_1$</td>
<td>211/292</td>
<td>72%</td>
<td>222/295</td>
</tr>
<tr>
<td>$T_2$</td>
<td>45/88</td>
<td>51%</td>
<td>78/107</td>
</tr>
<tr>
<td>$T_3$</td>
<td>40/106</td>
<td>37%</td>
<td>38/66</td>
</tr>
<tr>
<td>$T_4$</td>
<td>8%</td>
<td>N.P.</td>
<td>0%</td>
</tr>
</tbody>
</table>

salvage rate 1st type = successfully salvaged / all recurred cases
salvage rate 2nd type = successfully salvaged / operated for recurrence
N.P. = not published

Corrected and/or actuarial data for recurrence are seldom mentioned in the literature, nor have such data for salvage ever been described. This methodological shortcoming may partly be excused considering the small patient populations at risk. That also explains the broad range of recurrence rates and salvage rates (the latter are not summarized in Table 5). Therefore cumulative mean rates of recurrence rate and salvage rate may be more representative.

As could be expected, recurrence rates increase with increasing tumour size, both after primary radiotherapy and though to a lesser degree, after primary surgery. In general, recurrences after radiotherapy are more localized than after surgical failure. This may explain why salvage results are better after radiotherapy than after surgery.
Differences of up to 23% may occur depending on whether salvage is described as successful salvage among all recurred cases (1st type of salvage) or as successful salvage among all operated recurrences (2nd type of salvage). This difference can be ascribed to the number of patients with recurrent tumour who were not eligible for salvage. Reasons for non-eligibility can be numerous, but tend to increase with increased size of the original tumour, for both radiotherapeutic and surgical recurrences.

Radiotherapeutic treatment schedules and -techniques, when mentioned, showed relatively few differences between the different treatment centres, as can be read in Table I. More importantly, within each centre, treatment techniques may well have undergone a great evolution. However a breakdown of treatment results for these technical changes is seldom mentioned. Therefore, a breakdown of treatment results of radiotherapy in terms of different techniques could not be made.

In summary it can be said that the great diversity and incomplete description of statistical means and methods by which treatment results of glottic carcinoma are presented in the literature, make evaluation and comparison of surgical and radiotherapeutic treatments a difficult venture. These methodological shortcomings can partly be overcome if an author:

1. Describes actuarial results, both crude and corrected for intercurrent disease and those lost to follow up, and mentions the method applied (e.g. Kaplan-Meier, Berkson-Gage, Cutler-Ederer) [2,6,28].

2. Describes survival, disease-free survival (after primary treatment), recurrences, eligibility for salvage (and reasons for non-eligibility) and success of salvage (if possible of both types described above).

Results based on a literature review for the period 1975-1985 show equal survival for T1 and T2 glottic carcinoma, both after radiotherapy and after surgery. Scarce results for more extended primary tumours suggest the same. In the final choice between primary radiotherapy and primary surgery for glottic carcinoma, one should also consider the quality of life. Quality of life can be expressed in terms of conservation of normal swallowing, normal breathing and acceptance of complications especially concerning the quality of speech. Decision analytic methods can be used to evaluate if and how a definite choice may change if treatment results vary and if voice quality were taken into account [48].
Acknowledgement

The authors wish to thank Prof. J. Pliskin from the Ben Gurion University, Israël, for his critical comments.

6.5 References

1. American Joint Committee (AJC) for Cancer Staging and End Results Reporting. *Clinical staging systems for carcinomas of the larynx*. American Joint Committee (AJC) for Cancer Staging and End Results Reporting, Chicago 1978


7. Radiotherapy or surgery for T2N0M0 glottic carcinoma ?
A decision-analytic approach

7.1 Abstract

Decision analysis was used to evaluate the results of treatment of T2N0M0 glottic carcinoma as presented in the literature. Based on mean values for recurrence, salvage eligibility after recurrence and salvage success, the 5-year survival after radiotherapy and after surgery proved to be almost the same: 85% and 86%

Varying the recurrence rates and salvage rates demonstrated a marginal advantage for surgery in small tumours (T2a) and a major advantage in more extended (T2b) tumours if only survival is considered.

To take the quality of speech into account, a utility analysis was performed. A utility scale was defined ranging from 0.0 as the value for death to 1.0 for a successfully irradiated patient with preservation of normal speech. A utility value of 0.99 or less for the laryngectomized patient would favour radiotherapy over surgery for all T2 tumours. In patients with T2b tumours and in extreme circumstances, namely in institutions where failure rates of radiotherapy are extremely high and in those where recurrence rates after surgery are extremely low, an exact assessment of patient utilities may be pivotal. Under normal circumstances radiotherapy is preferred for T2N0M0 glottic carcinoma if both survival and the quality of speech are taken into account.

7.2 Introduction

According to the TNM-system for tumour classification the non-metastasized carcinomas of the vocal cord confined to the larynx with extension either to the supra- or subglottic regions with normal or impaired mobility, are classified as T2N0M0 glottic carcinomas [AJC, 1978, UICC, 1982]. Patients with a T2N0M0 glottic carcinoma can be treated with curative intention by:

(1). Primary surgery, either by conservative (partial) laryngectomy or by total laryngectomy.

---

1 L J A. Stalpers, MD, A L M. Verbeek, MD, W A J. van Daal, PhD, MD. 
(2). Primary radiotherapy with surgery if needed for loco-regional failure (so-called 'salvage surgery') [Million and Cassisi, 1984].

After primary treatment most patients with $T_2N_0M_0$ glottic carcinoma will be free of disease for the period under observation (e.g. 5 years disease-free survival). The major advantage of irradiation compared with laryngectomy or hemilaryngectomy is that the quality of voice is likely to be better and that a mutilating operation can be avoided [Lederman and Dalley, 1965].

Patients with tumour recurrence after primary treatment may undergo a salvage procedure. In general, salvage will consist of surgical removal of the recurrent process. Some patients with tumour recurrence after primary treatment are not eligible for salvage surgery for various reasons: A poor general condition, the presence of distant metastases, irresectability of the tumour or simply because patients refuse any further treatment. Spontaneous regression of recurrent tumours has never been described: These patients will die of the tumour. Patients eligible for salvage surgery may become free of disease again or the tumour will persist and they will die as a result.

The choice between primary radiotherapy and primary surgery in fact depends on treatment results expressed as disease-free survival, disease-free survival-after-recurrence and overall survival. In this study we have examined how decision analytic methods can be used in interpreting 5-year treatment results of the $T_2N_0M_0$ glottic carcinoma as presented in medical literature from 1974 - 1986. The questions are:

(1) Can we make a rational choice between radiotherapy and surgery as treatments for the $T_2N_0M_0$ glottic carcinoma?

(2) How does this choice between radiotherapy and surgery vary within the range of probabilities known from literature, e.g. recurrence rate, especially considering the different risks for patients with small ($T_{2a}$) and extended ($T_{2b}$) tumours?

(3) How does this choice alter if we could attach a value to the quality of voice after radiotherapy and after surgery? Or rather: What is it worth to the patient to have his voice preserved? Then, to what extent will the patient accept an increased rate of tumour recurrence? These more subjective aspects are difficult to quantify but can be of decisive importance in the choice patient and doctor have to make. Therefore, a choice merely based on survival, can shift by attaching a value to the quality of voice.
7.3 Literature Review

In a review of literature on treatment results of radiotherapy and surgery for non-metastasized glottic carcinoma, we looked for the treatment results expressed as disease-free survival and overall survival. Furthermore, we looked for indicators of treatment success such as the recurrence rate and the rate of salvage success [Stalpers et al., 1987]. Based on a series of well known clinical studies, the 5-year results for T2N0M0 glottic carcinoma are summarized in Table 7.1.

Table 7.1 • 5-year treatment results of T2N0M0 glottic carcinoma [30]

<table>
<thead>
<tr>
<th></th>
<th>Primary Radiotherapy</th>
<th>Primary Surgery</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Survival rate</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Crude</td>
<td>46%-72%</td>
<td>59%-72%</td>
<td>[5,12,19,24,32,36,5,19,29]</td>
</tr>
<tr>
<td>Adjusted</td>
<td>80%-91% mean=86%</td>
<td>88%</td>
<td>[7,9,12,14,25, 14]</td>
</tr>
<tr>
<td><strong>Recurrence rate</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All T2</td>
<td>16%-60% mean=30%</td>
<td>11%-31%</td>
<td>[3,4,5,7,12,14,16,19,22, 35,55, 2,3,5,14,19,29]</td>
</tr>
<tr>
<td>T2a</td>
<td>8%-67% mean=20%</td>
<td>11%</td>
<td>[14,22,38, 14]</td>
</tr>
<tr>
<td>T2b</td>
<td>27%-50% mean=30%</td>
<td>13%</td>
<td>[14,22,38, 14]</td>
</tr>
<tr>
<td><strong>Salvage rate</strong></td>
<td>50%</td>
<td>30%</td>
<td>[3,4,5,7,12,14,16, 19,22,25,32,2,3, 5,14,19,28,29]</td>
</tr>
</tbody>
</table>

Adjusted = for the follow-up adjusted 5-year survival

The rates have been structured and visualized in a decision tree presented in Figure 7.1. With this figure, 5-year survival was calculated (see Appendix). Based on mean values for recurrence rate and salvage success, Figure 7.1 gives an 85.0% calculated 5-year survival for primary radiotherapy. Compared with those articles...
which mention a for follow-up adjusted 5-year survival ranging from 80%-91%, the calculated survival represents a fair mean of the survival results explicitly mentioned. Analogously, the survival after primary surgery was calculated, showing an 86.0% 5-year survival rate. To our knowledge only one study, by Kaplan and Johns [1984] presented a comparable rate: A 5-year life table survival rate of 88.0%. The calculated survival for primary surgery (86.0%) is only slightly better than for primary radiotherapy (85.0%) and these values can be used as representative averages from the literature reviewed.

After primary radiotherapy tumour recurrence is high (30%) but the chance of being cured by salvage surgery is also fairly high (50%), resulting in 15% being free of disease after tumour recurrence. Tumour recurrence after surgery is lower (20%), but the chance of being salvaged after recurrence is also lower (30%), resulting in 6% free of disease following tumour recurrence.
Although the expected survival rates are almost equal for radiotherapy and surgery, we have to investigate the factors which may influence a decision. There are reasons why making a decision based on these figures could be hazardous. The conclusion of the decision tree analysis should be regarded with some reserve as our analysis is based on mean rates taken from the literature. For separate treatment centres and for particular populations different figures may be found which lead to different conclusions.

7.4 Sensitivity Analysis

In medical decision making analysis of the stability of a conclusion by varying rates in a decision tree is called a 'sensitivity analysis'. In the decision tree analysis we assumed all patients with T₂ tumours to be equally curable. In fact patients with T₂ glottic cancer are not one homogeneous group of patients. Therefore, based on the mobility of the vocal cords, a distinction was made between prognostic favourable T₂a tumours and prognostic unfavourable T₂b tumours. Although several reports have been published presenting radiotherapeutic results for T₂a and T₂b carcinomas separately [Harwood & De Boer, 1980; Kaplan et al., 1983; Notter et al., 1984; Van den Bogaert et al., 1983; Wang, 1974; Mills, 1979], we found only one paper describing comparable surgical results [Kaplan et al., 1983]. The recurrence rates and survival rates are summarized in table 7.1. Based on the decision tree approach, survival rates were calculated using the mean values for recurrence rates for patients with T₂a and T₂b tumours. In this analysis we presumed the salvage rate to be constant at 50% following radiotherapeutic failure and 30% following surgical failure. So, the calculated survival for patients with T₂a tumours is slightly higher after surgery than after radiotherapy, 92.3% and 90.0% respectively. The difference in the calculated survival is remarkably higher for T₂b tumours, being 90.9% following surgery and 85.0% following radiotherapy. Based on these calculated survivals, surgery should be preferred as primary treatment both for T₂a and particularly for T₂b tumours. However, the calculated survivals are only point estimates. If the recurrence rate after radiotherapy for T₂a tumours would be lower, for example 8% as found by Kaplan et al. [1983] instead of the mean value 20%, the survival from radiotherapy would be 96% instead of 90%. Then, compared with the 92% survival rate after surgery, radiotherapy would be preferred as primary
treatment for T2a tumours Figure 7.2 shows the survival rates for varying recurrence rates after radiotherapy.

Figure 7.2: Calculated survival rates for T2N0M0 glottic carcinoma following radiotherapy (solid line) or surgery (broken lines) for varying recurrence rates following radiotherapy. The results of T2a and T2b tumours are based on only one study presenting relatively low recurrence rates, this explains for the lower calculated survival rates for all T2 tumours based on six studies (see text).

The points where the solid radiotherapy line crosses the broken surgery lines are called the threshold-points. At these specific threshold-levels for the recurrence rates following radiotherapy, -being 15.4% for T2a, 18.2% for T2b and 28.0% for all T2 tumours- , the calculated survival rates are equal for radiotherapy and surgery. For lower recurrence rates following radiotherapy, radiotherapy will be preferred over surgery.

In figure 7.2, the recurrence rates following surgery were kept constant for T2a, T2b and all T2 tumours. The recurrence rates after surgery for T2a and T2b tumours were based on only one study [Kaplan et al, 1983] The surgical results of all T2 tumours, -based on six studies-, are generally worse than the one study by Kaplan et al. This finding may suggest that the results of surgery for T2a and particularly T2b tumours in most surgical clinics work out to be worse. Then, if we presume higher recurrence rates after surgery, the calculated survival after surgery
will be lower and the threshold-levels for the recurrence rate after radiotherapy will be higher.

The numerical relation between varying recurrence rates after surgery, as well as varying recurrence rates after radiotherapy and the consequently varying survival rates following surgery and radiotherapy, is displayed by the threshold-survival line in figure 7 3a. The threshold-survival line denotes the recurrence rates after surgery and the recurrence rate after radiotherapy for which the calculated survival rates are equal. Based on survival alone, radiotherapy will be preferred for recurrence rates left-above the threshold line and surgery will be preferred for recurrence rates right-under the threshold-line. For example, suppose we find a recurrence rate after radiotherapy for T2b tumours being 27%, as found by Wang [1974], then from figure 7 3 it can be read that radiotherapy will be preferred as primary treatment if we expect the recurrence rate after surgery to be equal or higher than 19.3%, as was found by some authors for all T2 tumours [Lippi et al., 1984, Skolnik et al., 1975]. Always presuming, of course, that the salvage rate after radiotherapy is 50% and after surgery 30%. Figure 7 3b and 3c give threshold-survival lines for some different salvage rates following radiotherapeutic failure (3b) and following surgical failure (3c).

7.5 Utility Analysis: The Quality of Speech

In medical decision making, utility analysis is based on the proposition that the preference for a treatment (or the utility of a treatment) depends on many aspects of the competing treatments, including life expectancy and quality of life. Utility assessment assumes that a loss in life expectancy can be compensated by a gain in quality of life, and that loss and gain can be measured.

In this particular clinical situation it means that in the treatment of glottic carcinoma, survival should not be the only parameter for treatment success. It seems obvious to consider the value the patient attaches to preservation of normal use of his/her voice in a choice between radiotherapy and surgery. Although radiotherapy may impair the quality of voice to some extent, [Stoicheff, 1975, Stoicheff et al., 1983] there is a general agreement that loss of speech after surgery, even after conservation surgery, is greater and more of a handicap [Harwood and Rawlinson, 1983].
Radiotherapy or surgery for

Figure 7.3a · *Survival thresholds for varying recurrence rates following radiotherapy and following surgery for T2N0M0 glottic carcinoma.*

Figure 7.3b · *Survival thresholds for varying recurrence rates following radiotherapy and following surgery for salvage rates following radiotherapeutic failure ranging from 20% to 80%.*

Figure 7.3c · *Survival thresholds for varying recurrence rates following radiotherapy and following surgery for salvage rates following surgical failure ranging from 0% to 60%.*
Several authors have recognized the problem of loss of laryngeal speech and the corresponding loss of quality of life [King et al, 1968, Minear and Lucente, 1979, Volin, 1980]. However, few attempts have been made to measure this loss [Harwood and Rawlinson, 1983, McNeil et al, 1981]. Such studies are essential since attaching a numerical value to quality of life could rationalize a decision to be taken.

In this analysis such a numerical value will be attached to different outcomes after treatment of $T_2N_0M_0$ glottic carcinoma by either surgery or radiotherapy. A utility scale is defined ranging from 0.0 as the value for death to 1.0 for a successfully irradiated patient with normal speech. Any qualitative loss by surgery, either as primary treatment or as salvage treatment, will be expressed as loss in quality of life compared with normal speech after successful radiotherapy.

At the extreme right of the decision tree (figure 7.1) the utilities $U$ for the quality of life have been attached to the various outcomes. In accordance with the calculation of survival, a weighted survival-value after both modalities of treatment can be calculated (see Appendix). Survival, adjusted for the utility of speech, is called expected utility [Weinstein and Fineberg, 1980]. Based on the data presented in figure 7.1, figure 7.4 gives expected utilities for an artificial speech following (partial) laryngectomy.

From figure 7.1 we can read that after radiotherapy, 70% of patients would have normal speech and only 15% would eventually have an artificial speech. After surgery, all patients will have (partial) laryngectomy and artificial speech. This explains why in figure 7.4 the expected utility or the ‘for speech quality corrected survival’ shows a higher sensitivity for variation of the utility of an artificial speech after surgery than after radiotherapy. For instance, a decrease in utility for the laryngectomized (artificial speech) from 0.8 to 0.7, leads to a decrease in expected utility after surgery from 0.69 to 0.60, while the one after radiotherapy decreases only from 0.82 to 0.81. It is striking that for a utility of 0.986 after laryngectomy the corresponding expected utilities for radiotherapy and surgery are the same. Or in other words: If a patient with $T_2N_0M_0$ glottic carcinoma states that on a utility scale between 0.0 for death and 1.0 for life with a natural speech, the quality of life after (partial) laryngectomy is equal or less than 0.986 (or 98.6%), radiotherapy is preferred over surgery as primary treatment. Always presuming of course, that the mean values for all $T_2$ tumours taken from the literature are representative. If, for instance, the recurrence rate after surgery was not 20% but 12%, as described by
Kaplan et al., the calculated survival would be 92% (i.e. much higher than the 85% survival after radiotherapy). Then patients with a lower utility for loss of voice due to surgery, would prefer surgery to radiotherapy.

![Expected utilities after radiotherapy and after surgery for T2N0M0 glottic cancer for varying utility values for the quality of voice after (partial) laryngectomy (=artificial speech).](image)

From the solid line in Figure 7.5a we can read that for a recurrence rate of 12% following surgery, the threshold-utility for artificial speech will be 0.92. Figure 7.5b gives a similar analysis on the variation of the recurrence rates after radiotherapy. It is questionable whether the quality of life value for the laryngetomized of 0.92 in situations with extremely high recurrence rates after radiotherapy or extraordinarily low recurrence rates after surgery for T2N0M0 glottic carcinoma are realistic. If so, radiotherapy is to be preferred as treatment for T2N0M0 glottic carcinoma.

### 7.6 Discussion and Conclusion

Rather simple decision analytic methods can be used as a tool in evaluating treatment results of the T2N0M0 glottic carcinoma as presented in literature, even though some figures are incomplete or subjective.
A preliminary calculation of survival yielded survival rates which were similar to the average survival rates described in the literature. But these survival rates were insufficient for us to make a founded choice between radiotherapy and surgery as a primary treatment for $T_2N_0M_0$ glottic carcinoma as treatment results vary between different treatment centres and for different populations of patients.

In a sensitivity analysis the influence of different recurrence rates and salvage rates was evaluated. Salvage of recurrent tumours after irradiation, unlike after surgery, does substantially increase survival figures for primary radiotherapy. This sensitivity analysis seems to support the following practical approach:

- If surgery is preferred as primary treatment for $T_2N_0M_0$ glottic carcinoma, it is essential to keep recurrence rates low. This might be accomplished by radical...
surgery. This may impair life qualities to an unacceptable level, thus outbalancing the reduction of recurrence rates. Little extra survival will be gained by improving the salvage rate.

- If we prefer, or want to improve, radiotherapy as a primary treatment for T2N0M0 glottic carcinoma, we first of all have to try to keep recurrence rates as low as possible. This might be accomplished by improved radiotherapeutic techniques, e.g., optimal localization and treatment planning [Fletcher et al., 1975, Peters and Fletcher, 1983]. Increasing the radiation dose does not seem to improve the results [Harwood and Tierie, 1979, Harwood and DeBoer, 1980, Harwood et al., 1981], but accelerated fractionation might [Withers et al., 1988].

But, in contrast to primary surgery, improvement of the salvage success after radiotherapeutic failure could significantly contribute to better survival results. This again stresses the importance of optimal tumour follow-up aiming at early detection and treatment of recurrences.

A utility analysis illustrated the effect on a decision by quantifying the quality of life of a laryngectomized patient compared to a successfully irradiated patient. Compared with death (utility = 0.0) and with radiotherapy (utility = 1.0), a minimal loss in the quality of life after (partial) laryngectomy compared with successful radiotherapy, will result in a preference for primary radiotherapy with surgical salvage in reserve for patients with T2N0M0 glottic carcinoma. This, if a decision was based on mean values for treatment results taken from the literature.

It depends on the representativity of these calculated utilities whether this preference for radiotherapy holds in (1) patients with more extended T2 tumours and in (2) treatment institutes where radiotherapeutic failure rates are extremely high or surgical recurrence rates are extraordinarily low. In these cases it would be useful to measure the utilities exactly [Stalpers et al., 1988].

Acknowledgement

The authors wish to thank Prof. J. Pliskin of the Ben Gurion University, Israel, and Prof. A.J. Van der Kogel, Nijmegen for their critical comments.
7.7 Appendix

Based on the recurrence rate ($P_{rec}$) and the salvage rate after recurrence ($P_{sal}$), formula (8.1) gives the calculated survival rate for the data used in the decision tree

$$P_{survival} = (1 - P_{rec}) + P_{rec} \times P_{sal} \quad \{1\}$$

For example, after surgery for T$_2$N$_0$M$_0$ glottic carcinoma we found a recurrence rate of 20% and a salvage rate of 30%, resulting in a calculated five-year survival rate

$$P_{survival} = (1 - 0.20) + 0.20 \times 0.30 = 0.86 \text{ or } 86\%$$

The for quality of speech corrected survival or 'expected utility' after radiotherapy ($E_{rad}$) can be calculated with formula (8.2)

$$E_{rad} = (1 - P_{rec}) \times U_{rad} + P_{rec} \times P_{sal} \times U_{sur} \quad \{2\}$$

where $U_{rad} = $ the utility after successful radiotherapy with natural speech and $U_{sur} = $ the utility after surgery with artificial speech. The expected utility after primary surgery, either conservative or total laryngectomy, can be calculated as

$$E_{sur} = (1 - P_{rec}) \times U_{sur} + P_{rec} \times P_{sal} \times U_{sur} \quad \{3\}$$

For example, for a recurrence rate after primary surgery of 20%, a salvage rate of 30% and a utility after (partial) laryngectomy of 0.92, the expected utility is 0.79 (or 79%).
7.8 References

1. American Joint Committee (AJC) for Cancer Staging and End Results Reporting. *Clinical staging systems for carcinomas of the larynx* American Joint Committee (AJC) for Cancer Staging and End Results Reporting, Chicago 1978


8. The role of yearly chest radiography in the early detection of lung cancer following oral cancer

8.1 Summary

In a study of 213 patients with oral cancer we investigated the incidence and prognosis of lung malignancies in patients offered a yearly chest radiography in the follow-up. Three conclusions can be drawn:

(1) Metastatic or primary lung cancer was diagnosed in 22 (10.3%) patients. The 2 year actuarial incidence rate of lung cancer following cancer of the oral cavity is 13%. No new lung cancers were detected after 2 years follow-up. This suggests that after this period, yearly chest radiography may be superfluous for the early detection of lung cancer.

(2) The survival rate of patients with a lung malignancy following cancer of the oral cavity is poor (1-year = 25%). The survival rate of patients detected by the yearly chest radiography without symptoms is higher than for patients detected after symptoms (p=0.006). It is not clear to what extent this different survival rate is biased by lead-time and selection of patients with a favourable prognosis. A randomized study would be required to assess whether patients with oral cancer do benefit from the yearly chest radiography compared with no regular chest radiography.

(3) From 22 patients with lung cancer, 13 (59%) were detected by chest radiography without symptoms. In the first year following oral cancer, 11 patients were diagnosed with lung cancer. Only 4 of these 11 patients (36%) were detected by chest radiography in an asymptomatic stage. The detection of patients with lung cancer in an asymptomatic stage may be increased by more frequent chest radiography examinations in the first year following oral cancer.

8.2 Introduction

In Western countries the annual incidence of squamous cell carcinoma of the oral cavity is 4:100,000 in men, and 1:100,000 in women. Cancer of the oral cavity is

---

predominantly a disease of the fifties [Waterhouse et al, 1982]. Smoking, abundant alcohol consumption, bad oral hygiene and a combination of these are the most important etiologic factors [Wynder et al. 1957; Vogler et al., 1962; Rothman and Keller, 1972; Stockwell and Lyman, 1986]. Both radiotherapy and surgical resection are successfully applied as primary treatment [Mendenhall et al., 1981; Million and Cassisi, 1984; Zagars and Norante, 1983]. However, the lifespan of patients will be limited by metastases and second primary malignancies. Increased risks for the development of second primary malignancies following another cancer have been described as early as 1932 by Warren and Gates. Table 8.1 gives an overview of the literature on incidences of second primary malignancies following oral cancer.

Smoking and alcohol abuse are related to the development of second primary malignancies [Wynder et al, 1977; Batasakis, 1979]. The etiologic role of immunologic and genetic factors in the development of multiple primary malignancies is not clear [De Vries et al., 1986; De Vries et al., 1987]. Early detection and early treatment of these second primary malignancies and metastases may improve the survival of patients with oral cancer. This is the rationale of regular physical and radiological lung examination in the follow-up of patients with cancer of the oral cavity.

In this study we looked for: (1) The incidence of pulmonary malignancies, both primary and metastatic, following oral carcinoma, and (2) the value of the yearly chest radiography in the early detection of these pulmonary malignancies.

8.3 Patients and Methods

Between January 1st 1979 and January 1st 1986, 238 patients with a squamous cell carcinoma of the oral cavity were admitted for treatment to the University Hospital St. Radboud in Nijmegen, The Netherlands. Eligible for this study were all patients admitted for primary treatment of carcinoma of the oral cavity. Excluded were 25 patients who came for a recurrency treatment. Included were eight patients cured from another malignancy before occurrence of the carcinoma of the oral cavity. Also included were four patients with synchronous oral cancer and lung cancer. Eventually the medical records of 213 patients were reviewed for this study.

For the study of the lung cancer incidence rate, follow-up closed on January 1st 1987. The median observation period is 49 months. In patients with lung cancer
following oral cancer, follow-up closed on July 1st 1988, ensuring at least 18 months follow-up of lung cancer.

Table 8.1 • Literature review of incidence ratios of second primary malignancies following cancer of the oral cavity

<table>
<thead>
<tr>
<th>Reference</th>
<th>Period</th>
<th>Patients at risk</th>
<th>Second primary malignancies</th>
</tr>
</thead>
<tbody>
<tr>
<td>BERG et al., 1970</td>
<td>'49-'62</td>
<td>3443</td>
<td>248 (7.1%)</td>
</tr>
<tr>
<td>COHN &amp; PEPPARD, 1980</td>
<td>'70-'79</td>
<td>264</td>
<td>44 (16.7%)</td>
</tr>
<tr>
<td>GLUCKMAN et al., 1980</td>
<td>'77-'79 (I)</td>
<td>72</td>
<td>9 (12.5%)</td>
</tr>
<tr>
<td></td>
<td>'74-'78 (II)</td>
<td>158</td>
<td>28 (17.7%)</td>
</tr>
<tr>
<td></td>
<td>'50-'79 (III)</td>
<td>1551</td>
<td>187 (12.0%)</td>
</tr>
<tr>
<td>TEPPERMANN &amp; FITZPATRICK, 1981</td>
<td>'58-'75</td>
<td>377</td>
<td>101 (26.8%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>82 (21.8%)</td>
</tr>
<tr>
<td>HORDIJK &amp; DE JONG, 1983</td>
<td>58-'79</td>
<td>152</td>
<td>26 (17.1%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>22 (14.4%)</td>
</tr>
<tr>
<td>BLACK et al., 1983</td>
<td>'50-'78</td>
<td>1551</td>
<td>287 (18.5%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>135 (8.7%)</td>
</tr>
<tr>
<td>DE VRIES et al., 1986</td>
<td>'63-'84</td>
<td>210</td>
<td>38 (18.1%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>23 (11.0%)</td>
</tr>
<tr>
<td>YELLIN et al., 1986</td>
<td>'43-'83</td>
<td>372</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>11 (3.0%)</td>
</tr>
<tr>
<td>Present study, 1988</td>
<td>'79-'87</td>
<td>213</td>
<td>29 (13.6%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>23 (10.8%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5 (2.3%)</td>
</tr>
</tbody>
</table>

Table 8.2 presents the distribution of patients by the site of the primary tumour and the actuarial five-year survival rate. At diagnosis of the oral cavity cancer, the sex-ratio was 151 men and 62 women = 2.4 ; the mean and median age was 62.5 years (range 27-91 years). All patients were classified and staged by the UICC-TNM norms for tumour classification [UICC, 1980]. Surgery was performed on operable patients. Radiotherapy was given (1) pre-operatively up to 40 Gy in T3 and T4 carcinomas (before 1983), (2) post-operatively up to 60-64 Gy if surgery was not radical or if tumour-spill was likely or in the presence of extranodal spread and (3)
in inoperable patients. Chemotherapy was reserved for palliative care of incurable carcinomas.

Table 8.2 · *Tumour site and actuarial five-year survival of 213 patients with cancer of the oral cavity*

<table>
<thead>
<tr>
<th>Tumour site</th>
<th>Number</th>
<th>Five-year Survival</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tongue</td>
<td>66</td>
<td>49%</td>
</tr>
<tr>
<td>Floor of mouth</td>
<td>59</td>
<td>60%</td>
</tr>
<tr>
<td>Buccal mucosa</td>
<td>27</td>
<td>43%</td>
</tr>
<tr>
<td>Retromolar trigone</td>
<td>15</td>
<td>70%</td>
</tr>
<tr>
<td>Alveolar ridge</td>
<td>34</td>
<td>52%</td>
</tr>
<tr>
<td>Lip</td>
<td>12</td>
<td>76%</td>
</tr>
<tr>
<td>All sites</td>
<td>213</td>
<td>52%</td>
</tr>
</tbody>
</table>

Table 8.3 presents the follow-up protocol for patients treated for oral cancer. The follow-up includes a yearly chest radiography consisting of an antero-posterior view and a lateral view.

Table 8.3 · *Protocol for the follow-up of patients with oral cavity carcinoma (Nijmegen, The Netherlands)*

I. Policlinical examinations
   1st year     bimonthly
   2nd year     threemonthly
   3rd year     fourmonthly
   4th year     semiannual
   >4 years     yearly

II. Yearly X-chest examination
   - antero-posterior view
   - lateral view

III. Blood- and/or urine-analysis on indication
The pre-operative chest radiogram used for the staging of oral cancer is defined as the first yearly chest radiogram. All chest radiograms were reviewed at least three times before a definite report was given: A first examination by a resident in diagnostic radiology, review by a radiologist and finally discussion in the weekly meeting between a diagnostic radiologist and a radiotherapist. If the chest radiography was suspect for a pulmonary malignancy, further examinations were performed, generally including pulmonary planigraphy. To plan the further diagnostic and therapeutic strategy, each patient was evaluated by the interdepartmental Workgroup for Lung Cancer. Further histologic diagnosis depended on the feasibility of invasive diagnostic techniques (bronchoscopy, cytologic puncture or diagnostic thoracotomy) and on the curability of the patient. Surgery was performed on curable and operable lung cancer; radiotherapy and chemotherapy were reserved for palliative care.

In 1932, Wareen and Gates defined three diagnostic criteria for second primary malignancies. We applied the three criteria as modified by Gluckman et al. in 1980:

1. The neoplasms must be clearly malignant as determined by histologic evaluation.
2. Each neoplasm must be geographically separate and distinct. The lesions should be separated by normal-appearing mucosa. If a second neoplasm is contiguous to the initial primary tumour or is separated by mucosa with intraepithelial neoplastic change, the two should be considered as confluent growths rather than multicentric carcinomas.
3. The possibility that the second neoplasm represents a metastasis should be excluded. The observation that the invasive carcinoma arises from an overlying epithelium that demonstrates a transition of carcinoma in situ to invasive carcinoma is helpful, and when the separate foci have significant differences in histology, the diagnosis of separate primary cancers is appropriate.

Patients with no histologically proven second primary malignancy in the lung, are defined as patients with metastatic lung cancer. Patients with a radiologic progressive mass and with clinical deterioration, but without histological confirmation of a malignancy, are also defined as patients with metastatic lung cancer. In general these patients were in too bad a condition to allow histologic confirmation or curative surgery.
Patients with a radiographic lesion suspect for malignancy which did not show radiographic or clinical progression within a year, were not considered as patients with a pulmonary malignancy.

The incidence of a second primary malignancy or a pulmonary metastasis is calculated as a crude rate for all patients and as an actuarial incidence rate adjusted for the duration of follow-up by the Kaplan-Meier method [Kaplan and Meier, 1958, Peto et al., 1977]

The survival of patients with a pulmonary malignancy was calculated by the Kaplan-Meier method; the survival curve of patients detected by the yearly chest radiography without symptoms and the survival curve of patients detected after symptoms are compared with the logrank-test [Kaplan and Meier, 1958, Cutler and Ederer, 1958]

8.4 Results

(a) Incidence

\[\begin{array}{c|c}
\text{Year} & \text{Incidence Rate} \\
\hline
0 & 0 \\
1 & 2 \\
2 & 5 \\
3 & 10 \\
4 & 20 \\
5 & 30 \\
6 & 40 \\
7 & 50 \\
\end{array}\]

\[\text{Figure 8.1} \cdot \text{Actuarial incidence rate of metastatic or primary lung cancer following oral cancer (N=22/213), the figures indicate the number of patients at risk, dotted lines indicate 95\% confidence limits.}\]

During the follow-up period of 213 patients with squamous cell carcinoma of the oral cavity, 29 (13.6\%) had a second primary malignancy. Twenty-three patients (10.8\%) had a localization in the upper aerodigestive tract, of which five (2.3\%)
were localized in the lung. Of these five histologically proven second primary malignancies in the lung, two were adenocarcinomas and three were squamous cell carcinomas.

Another 17 patients had metastatic lung cancers, resulting in 22 patients (10.3%) with a pulmonary malignancy.

Figure 8.1 shows the actuarial incidence rate of pulmonary malignancies, both primary and metastatic: 8.5% within one year following oral cancer and 13.0% within two years. After more than two years following cancer of the oral cavity, no more primary lung cancers or lung metastases were seen in 100 patients at risk.

In 13 out of 22 patients (59%) with a pulmonary malignancy, the tumour was detected without symptoms by the yearly chest radiography. Eleven patients had a lung malignancy within one year following treatment for oral cancer. Only four of these 11 patients (36%) were detected at the beginning of that year without symptoms by the first chest radiography.

(b) Survival

![Survival rates of 22 patients with a pulmonary malignancy, without symptoms (solid line, n=13) and with symptoms (broken line, n=9) at diagnosis.](image)

The survival curve of 22 patients with a pulmonary malignancy following oral cancer in Figure 8.2 shows a meager one-year survival of 25%. Figure 8.2 shows a significantly worse survival curve in nine patients with a pulmonary malignancy diagnosed after symptoms compared with 13 patients detected by yearly chest
Early Detection of Lung Cancer

radiography without symptoms (p=0.006). Median survivals are 2.0 months and 6.8 months respectively. All patients detected with a pulmonary malignancy after symptoms died within 18 months. Four out of 13 patients (30%) with a pulmonary malignancy detected by yearly chest radiography without symptoms were still alive after 18 months follow-up.

Only one out of five patients with a histologically confirmed primary lung cancer survived more than 18 months. In this specific patient, the lung tumour was detected synchronously with the oral cancer by a chest radiography. The patient died three years later of a local recurrence of the oral tumour.

8.5 Discussion

(a) Incidence
Table 8.1 presents some incidence rates of second primary malignancies following cancer of the oral cavity as described in the literature; incidence rates range from 7.1% to 26.8%. In our study the incidence rate is 13.6%.

Corresponding with the literature referred in table 8.1, most of these second primary tumours (23/213 = 10.8%) are localized in the upper aerodigestive tract. Relatively few second primary malignancies were localized in the lung (5/213 = 2.4%). This figure corresponds with the 3.0% incidence of lung cancer as found by Yellin et al. in 1986. The relatively low incidence of second primary malignancies in the lung may be partly due to the strictness of the criteria by Warren and Gates: In medical practice it may sometimes be extremely difficult or unethical to obtain representative histologic specimens from critically ill patients just to allow differentiation between metastatic and primary lung cancer. For this reason, some tumours defined as metastatic cancers may in fact have been primary lung malignancies [Adkins et al., 1968; Cahan, 1977; Lefor et al., 1986]. The distinction between metastatic and primary lung cancer can be of prognostic significance. However, in clinical practise it is frequently not feasible to make such a distinction properly. Therefore we evaluated the efficacy of the yearly chest radiography for both types of malignancies together.

The 2 years actuarial incidence rate of pulmonary malignancies following carcinoma of the oral cavity rises up to 13%. We found no more pulmonary malignancies after two years follow-up. This finding suggests that the yearly chest radiography becomes superfluous after two years following a carcinoma of the oral
cavity. Our findings should be regarded with some reserve as figures for comparison are lacking in the literature: None of the articles reviewed used the actuarial or life-time estimate of the lung cancer incidence rate [Kaplan and Meier, 1958; Cutler and Ederer, 1958; Peto et al., 1977]. The feature that lung cancer risk is only enhanced for a limited time following oral cancer might therefore have gone unnoticed in former studies.

Most pulmonary malignancies (59%) have been detected by a yearly chest radiography in an asymptomatic stage. However, in the first year, only 36% of the lung malignancies were detected without symptoms by the first chest radiography. Apart from statistical chance, a biological and a technical explanation may be given for the low detection rate in the first year. The biological explanation is that relatively fast growing lung malignancies are preponderant in the first year compared to following years. The technical explanation is the lack of a former chest radiography in the first year for comparison with the first one; repeated sequential chest radiography may be essential for recognition of radiographic chances suspect for malignancy. If so, malignant pulmonary changes will be detected earlier without symptoms by more frequent chest radiography examinations in the first year following oral cancer. Comparative figures from the literature are not known to us.

(b) Survival
Cancer of the lung has a poor prognosis with 1-year survival rates ranging from 25 to 35% [Fontana et al., 1984; Nou, 1984; Ries et al., 1983]. The value of a regular chest radiography for population screening for lung cancer has been discredited by the poor improvement, if any, in the prognosis of lung cancer [Bailar, 1984]. Although screening for lung cancer may be inefficient for the general population, a regular chest radiography may be more effective for patients with an enhanced risk in order to develop a pulmonary malignancy, either primary or metastatic. This study demonstrates a poor prognosis in patients with lung cancer following cancer of the oral cavity (1 year = 22%; 2 year = 15%). The survival rate of patients with a lung malignancy detected by the yearly chest radiography was significantly higher than of patients with symptoms, despite small population numbers and a small difference in median survival of only 4.5 months. However, it is not clear how far this higher survival rate in asymptomatic patients is biased by:
(1) The lead-time between asymptomatic detection by the yearly chest radiography and the time that a patient would present with symptoms (see appendix) and

(2) the selection of relatively benign, slowly growing tumours in patients detected by a regular chest radiography compared to aggressive, fast growing tumours in patients with symptoms [Feinleib and Zelen, 1969] The exact benefit of regular chest radiography in the follow-up of cancer of the oral cavity can only be established by a randomized trial of early diagnostic manoeuvres [Sackett et al., 1985] However, as without lead-time correction the differences in survival between symptomatic and asymptomatic patients is only a matter of months, the benefit of a regular chest radiography may in fact be small

8.6 Conclusions

Based on this study on the relevance of the chest radiography in the follow-up of 213 patients with carcinoma of the oral cavity, three conclusions can be drawn

(1) An increased occurrence of lung cancer following oral cancer seems to be restricted to the first two years of clinical follow-up Hence, the application of yearly chest radiography may be limited to this short period

(2) A prognostic benefit of chest radiography cannot be excluded, but the benefit will probably only be marginal To assess this benefit, a randomized trial would be necessary comparing yearly chest radiography with no chest radiography

(3) More frequent chest radiography in the first year following treatment for oral carcinoma, may enhance the potency of a regular chest radiography to detect a pulmonary malignancy in an asymptomatic stage

8.7 Appendix

Lead-time bias is a well-known pitfall in a cancer screening program where an increased survival time after diagnosis does not necessarily mean that early diagnosis was helpful It might simply mean that the time of diagnosis has been advanced without necessarily meaning that the time of death has been delayed [Feinleib and Zelen, 1969]
Figure 8.3: Graph representing a hypothetical natural course of lung cancer from onset towards the time that cancer can be diagnosed by chest radiography, towards the time that a tumour gives rise to symptoms and eventually evolving towards the time of death from a fatal volume.

For example, figure 8.3 displays the hypothetical course of disease in a patient with lung cancer. Let us presume that tumour growth is such that it will give rise to symptoms after three years and evolve into death after four years. Survival time is said to be one year. With chest radiography we might have diagnosed the tumour a year earlier in an asymptomatic stage. The time between early detection and the time of onset of symptoms is called the lead-time. Simply by advancing the time of diagnosis it seems as if we have prolonged survival time from 1 to 2 years. In fact, we did not change the natural course of disease and the patient dies four years following onset of tumour growth. Because of lead-time, the survival of patients diagnosed without symptoms seems better than the survival of patients detected with symptoms.

For instance, in Figure 8.2, if we presume an average lead-time of only four months, the survival curve of patients 'detected without symptoms' would shift about four months to the left and would converge with the curve of patients 'diagnosed after symptoms', and the previous observed difference between the two survival curves would disappear.
8.8 Acknowledgement

The patients described in this study were treated and followed by the Head and Neck Oncology Group and the Lung Cancer Group of the St Radboud University Hospital, Nijmegen, The Netherlands, wherein the following departments participated:

- Otorhinolaryngology (head: Prof. Dr. P. van den Broek),
- Maxillofacial Surgery (head: Prof. Dr. H.P.M. Freihofer),
- Medical Oncology (head: Prof. Dr. D.J.Th. Wagener),
- Diagnostic Radiology (head: Prof. Dr. J.H.J. Ruys),
- Pulmonology (head: Prof. Dr. C.L.A. Van Herwaarden) and
- Radiotherapy (head: Prof. Dr. W.A.J. van Daal).

8.9 References


9. Yearly chest radiography in the early detection of lung cancer following laryngeal cancer

9.1 Summary

In a retrospective study of 556 patients (505 men, 51 women) with laryngeal cancer the incidence and prognosis of lung malignancies was studied in patients who were yearly examined by chest radiography. In 69 patients (12.4%) a lung malignancy was diagnosed of whom 28 patients with a histologically confirmed second primary malignancy. All 69 patients were men. The incidence of radiological assessed lung malignancies, both second primary and metastatic lung cancer, is higher and more prolonged following supraglottic carcinoma then following glottic carcinoma. In 47 patients (68%) without symptoms, the lung malignancy was detected by yearly chest radiography. The survival rate in patients with lung cancer detected by the yearly chest radiography was significantly higher than in patients diagnosed after symptoms (median survival = 10 and 5 months respectively). However, because of a lead-time between early radiologic diagnosis of lung cancer and the time a tumour would be diagnosed following symptoms, the observed survival benefit of yearly radiography is in fact much less or even nihil.

9.2 Introduction

In Western countries the incidence of laryngeal cancer is 114 in 1,000,000 in men and 6 in 1,000,000 in women [Waterhouse et al, 1982]. It predominantly is a disease of the sixth decade [Zagars and Norante, 1982].

Laryngeal cancer may give relatively early rise to symptoms, especially persistant hoarseness, and curative treatment can be relatively well attained both by radiotherapy and by surgery [Stalpers et al, 1987]. As early as 1932, Warren and Gates described an increased association between cancer in the head and neck region and second primary tumours [Warren and Gates, 1932]. Second primary tumours are predominantly seen in the lung, occuring in 2% to 12% of patients [Martin et al., 1979; Hordijk and De Jong, 1983]. This increased incidence of lung

---

   Accepted for publication in Cancer
cancer following laryngeal cancer is probably due to cigarette-smoking as a common etiologic factor [Wynder et al., 1977] With improving results of the local treatment of laryngeal cancer, secondary tumours become a main threat to the lifespan of patients Early detection and early treatment of second primary malignancies may eventually improve the overall prognosis of these patients [Hordijk and De Jong, 1983, De Vries et al., 1986] This is the rationale of regular chest radiography in the follow-up of patients with laryngeal cancer

In this study we looked for (1) The incidence of second primary malignancies in the lung and lung metastasis in patients treated for laryngeal cancer and (2) the role of yearly chest radiography in the early detection of these lung malignancies

### 9.3 Patients and Methods

From January 1979 until January 1986, 590 patients with a squamous cell carcinoma of the larynx were admitted for treatment in the St Radboud University Hospital, Nijmegen, The Netherlands

Eligible for this study were all patients admitted for primary treatment Excluded were 34 patients who came for treatment of a recurrence Eventually the medical records of 556 patients were reviewed for this study For estimation of the lung cancer incidence, follow-up closed in July 1986 The median observation duration was 37 months For estimation of the lung cancer survival rate, follow-up closed in July 1989, ensuring at least a follow-up of three years The mean follow-up was 52
months. The male:female ratio of the 556 patients was 9.9 (505 men, 51 women). The median age at diagnosis of the laryngeal tumour was 61 years (range: 32-96 years). Figure 9.1 shows the distribution of sex and age.

All tumours were staged according to the TNM-classification for tumours by site of the tumour (table 9.1).

---

Table 9.1 • Distribution of 556 patients primarily treated for laryngeal cancer by tumour site and stage [AJC, 1978; UICC, 1978].

<table>
<thead>
<tr>
<th>Site</th>
<th>Stage:</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supraglottic</td>
<td></td>
<td>35</td>
<td>42</td>
<td>58</td>
<td>106</td>
<td>241</td>
</tr>
<tr>
<td>Glottic</td>
<td></td>
<td>173</td>
<td>62</td>
<td>58</td>
<td>23</td>
<td>309</td>
</tr>
<tr>
<td>Subglottic</td>
<td></td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>208</td>
<td>105</td>
<td>110</td>
<td>133</td>
<td>556</td>
</tr>
</tbody>
</table>

Treatment consisted of radiotherapy in T₁ and T₂ carcinomas (70-74 Gy in daily fractions of 2.0 Gy in 6-8 weeks). Surgical salvage was reserved for recurrent tumours. In more advanced stages, T₃ and T₄ primary surgery was performed with or without pre-operative radiotherapy (25.0 Gy in 5 daily fractions of 5.0 Gy in the week preceding operation) [Kazem et al., 1982]. Table 9.2 gives the follow-up protocol of patients with laryngeal cancer.

The follow-up includes a yearly chest radiography consisting of an postero-anterior and a lateral view. All chest radiograms were reviewed at least threefold before a definite report was given, both by a resident radiologist and a staff radiologist and finally by a radiologist and the prescribing oncologist. If a chest radiogram was suspected for malignancy, further radiologic examinations were performed. Depending on the feasibility of invasive diagnostics and on the curability of the patient, bronchoscopy and/or mediastinoscopy were performed. Surgery was performed in curable and operable lung cancer; radiotherapy and chemotherapy were reserved for palliative care.
Table 9.2 · Protocol for the follow-up of patients with laryngeal cancer

<table>
<thead>
<tr>
<th>I. Clinical examination</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st year</td>
</tr>
<tr>
<td>2nd year</td>
</tr>
<tr>
<td>3rd year</td>
</tr>
<tr>
<td>4th year</td>
</tr>
<tr>
<td>&gt;4 years</td>
</tr>
</tbody>
</table>

II. Yearly chest radiography
- Antero-posterior view
- Lateral view

III. Blood- and/or urine-analysis
on indication

Second primary lung cancer was defined by the three criteria of Warren and Gates as modified by Gluckman et al. in 1980 [Gluckman et al., 1980]:

(1) The neoplasms must be clearly malignant as determined by histologic evaluation.

(2) Each neoplasm must be geographically separate and distinct. The lesion should be separated by normal-appearing mucosa. If a second neoplasm is contiguous to the initial primary tumor or is separated by mucosa with intraepithelial neoplastic change, the two should be considered as confluent growths rather than multicentric carcinomas.

(3) The possibility that the second neoplasm represents a metastasis should be excluded. The observation that the invasive carcinoma arises from an overlying epithelium demonstrating a transition of carcinoma in-situ to invasive carcinoma is helpful, and when the separate foci have significant differences in histology the diagnosis of separate primary cancers is appropriate.

Patients with no histologically proven second primary in the lung are defined as patients with metastatic lung cancer. In this study patients with a radiologic progressive mass and clinical deterioration, but without histologic confirmation of a malignancy, are also defined as patients with metastatic lung cancer. In general
these patients were in too bad a condition to allow histologic confirmation or curative surgery

Patients with a radiographic lesion suspect for a malignancy which did not show radiographic or clinical progression within a year, were identified as having no pulmonary malignancy

The incidence of a second primary malignancy or a pulmonary metastasis was calculated as a crude rate of all patients and as an actuarial incidence rate corrected for the duration of follow-up [Cutler and Ederer, 1958, Peto et al, 1977]

The survival of patients with a pulmonary malignancy was estimated by an actuarial method, the survival curve of patients detected by the yearly chest radiography without symptoms and the survival curve of patients detected after symptoms were compared with the logrank-test [Cutler and Ederer, 1958]

9.4 Results

In 69 of 556 patients (12.4%) a malignancy of the lung following laryngeal cancer was diagnosed. Two patients with a histologically confirmed metastasis from breast cancer were not included in this number, all 69 patients were men. In 28 patients (5.0%) the tumour was histologically confirmed as a second primary tumour, in 11 patients (2.0%) the tumour was a histologically confirmed metastasis. In 30 patients (5.4%) no histological confirmation could be obtained for a progressive lung mass and were hence defined as having metastatic lung cancer. Figure 9.2 shows the actuarial incidence rate for a lung malignancy, being 17% after 5 years and increasing to more than 20% after 7 years.
A lung malignancy was seen more frequently following supraglottic cancer (41/241 = 17.0%) than following glottic cancer (28/309 = 9.1%). The difference was highly significant (p=0.001). Figure 9.3 shows the actuarial incidence rate of lung cancer, stratified by site of the laryngeal tumour. The incidence rate of lung cancer following supraglottic cancer is 17% after 5 years rising up to 32% after 7 years. The curve of the incidence rate following glottic cancer slopes down after two years follow-up (9%) up to 13% after five years. After five years, no more lung malignancies were seen in 67 patients with glottic cancer.

Figure 9.3 - Incidence rates of lung cancer following cancer of the glottic or supraglottic larynx (patient numbers at risk are indicated)

Figure 9.4 - Survival curves of patients with lung cancer following laryngeal cancer
Of 69 patients with a lung malignancy, 47 (68%) had no symptoms and were only detected by yearly chest radiography. The tumours in the other 22 patients (32%) were diagnosed after they had lead to symptoms between the yearly radiologic examinations. Figure 9.4 shows the survival curve of all 69 patients with a lung malignancy and the survival curves after stratification by mode of presentation. The survival of patients in whom the tumour was detected without symptoms is significantly better than of patients with symptoms (p<0.05), the median survival being 10 and 4 months respectively.

Figure 9.5 gives the survival curves after stratification by lung cancer treatment. Patients who had surgery did significantly better than patients who had radiotherapy or no treatment at all (p<0.05); the median survivals are 26, 11 and 3 months respectively.

9.5 Discussion and Conclusion

INCIDENCE: As described in table 9.3, the incidence of lung cancer following laryngeal cancer and lung cancer has been reported to range from 1.1% to 11.7%. This wide range may be explained both by different length of follow-up periods and probably more important, by different criteria for diagnosis of a second primary malignancy [Cahan, 1977; Moertel, 1977].
To minimize the latter confusion, several authors stress the need of a histologic verification of the secondary lung tumour [Cahan, 1977, Moertel 1977, Gluckman et al, 1980] The requirement of histological confirmation of second primary malignancies may result in a selection of 'pure' second primaries but may result in an underestimation of the real incidence rate Moreover, overemphasis on the need for a histologically distinction between metastatic and primary lung cancer may distract from the clinical purpose of the yearly chest radiography, namely to detect a lung malignancy in a curable stage Therefore, we evaluated the role of yearly chest radiography for metastatic and primary lung cancer together Then we see that the risk of developing lung cancer following supraglottic carcinoma is higher than following glottic cancer This finding agrees with other studies by Wagenfeld et al (1980, 1981)

<table>
<thead>
<tr>
<th>Author</th>
<th>Period</th>
<th>Patients</th>
<th>Second primary malignancies</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>All sites (%)</td>
</tr>
<tr>
<td>Sherman et al, 1967</td>
<td>1955-65</td>
<td>187</td>
<td>16.0</td>
</tr>
<tr>
<td>Goffinet et al, 1973</td>
<td>1957-70</td>
<td>535</td>
<td>16.0</td>
</tr>
<tr>
<td>Brown, 1978</td>
<td>1958-74</td>
<td>1,600</td>
<td>16.0</td>
</tr>
<tr>
<td>Martin et al, 1979</td>
<td>1960-77</td>
<td>496</td>
<td>16.0</td>
</tr>
<tr>
<td>Gluckman et al, 1980</td>
<td>1974-78</td>
<td>189</td>
<td>17.9</td>
</tr>
<tr>
<td>Wagenfeld, 1980, 1981</td>
<td>1965-74</td>
<td>903</td>
<td>7.5</td>
</tr>
<tr>
<td>Hordijk et al, 1983</td>
<td>1958-79</td>
<td>691</td>
<td>20.5</td>
</tr>
<tr>
<td>Rodriguez et al, 1984</td>
<td>1979-82</td>
<td>286</td>
<td>9.9</td>
</tr>
<tr>
<td>Miyahara et al, 1985</td>
<td>1958-81</td>
<td>1,389</td>
<td>9.9</td>
</tr>
<tr>
<td>De Vries et al, 1986</td>
<td>1963-84</td>
<td>748</td>
<td>13.9</td>
</tr>
<tr>
<td>Present study, 1989</td>
<td>1979-86</td>
<td>556</td>
<td>7.7</td>
</tr>
</tbody>
</table>

*excluding metastatic lung cancer

Moreover we found that while an increased lung cancer risk persists for many years following supraglottic carcinoma, the incidence rate slopes down after 2 years following glottic carcinoma and becomes asymptotic after less than 5 years A similar pattern, although with lower incidence rates, was found if we confined the population to the 28 patients with a histologically confirmed second primary lung
cancer (not presented here) These findings confirm the hypothesis of Wagenfeld et al that different mechanisms underlie the tumour development in patients with supraglottic and glottic carcinoma, despite common etiologic factors.

Second primary lung cancer following laryngeal cancer has been described as a feature predominantly seen in males [Sherman et al., 1967, Brown, 1978]. We found no lung malignancy associated with laryngeal cancer at all in 51 women, while 7 cases (13%) were expected. The finding could highly unlikely be attributed to chance alone (p=0.01).

SURVIVAL The survival of patients with a lung malignancy following laryngeal cancer is poor [Sherman et al., 1967, Yellin et al., 1986] and similar to the lung cancer survival in the general population [Ries et al., 1983, Nou, 1984].

The higher survival rate of patients detected by yearly chest radiography compared with patients diagnosed after symptoms suggests a benefit of yearly chest radiography. However, median survivals differ only 6 months. Considering a lead-time between early diagnosis of lung cancer and the time a tumour would be diagnosed following symptoms, the observed survival benefit of yearly chest radiography is in fact much less [Feinleib and Zelen, 1969]. If we would correct the survival of patients detected by chest radiography for only several months lead-time, the survival curve would converge with the curve of patients diagnosed after symptoms, and the earlier observed difference tends to disappear. However, the fact that almost all patients diagnosed following symptoms have died within two years, while almost 20% of patients “without symptoms” show a long term survival, suggests a probably small benefit of yearly chest radiography.

The finding that yearly chest radiography hardly contributes to an improved survival of patients with lung cancer (and thus with laryngeal cancer) agrees with the disappointing results of population screening for lung cancer [Bailar, 1984, Fontana et al., 1984].

Although patients who had surgery for lung cancer show a better survival than those who had not, the benefit of surgery is probably spurious. The policy in our clinic is to reserve surgery for curation. Therefore, patients with a poor prognosis due to a poor general health or to large or multiple lung tumours are not eligible for surgery.
Especially for patients with a prolonged enhanced risk of developing lung cancer we have to intensify the search for (a) better early detection methods (b) the underlying causes and (c) prevention of secondary cancers

CONCLUSIONS
(1) The incidence of lung cancer is both higher and more prolonged in the follow-up of patients with supraglottic carcinoma than in patients with glottic carcinoma
(2) An enhanced risk to develop lung cancer following laryngeal cancer is probably an exclusive feature in men
(3) The survival of patients with lung cancer detected by yearly chest radiography is better than in patients diagnosed after symptoms. The prognostic benefit of chest radiography could disappear after correction for lead-time

96 References

1 AJC - American Joint Committee for Cancer Staging and End Results Reporting. *Clinical Staging System for Carcinoma of the Larynx*. American Joint Committee (AJC) for Cancer Staging and End Results Reporting, Chicago 1978


3 Brown M Second primaries in cases of the larynx *J Laryngol Otol* 92 991-996, 1978

4 Cahan WG Montemayor PB Cancer of the larynx and lung in the same patient *J Thorac Cardiovasc Surg* 44 309-320, 1962


6 De Vries N, Van der Waal I, Snow GB Multiple primary tumours in laryngeal cancer *J Laryngol Otol* 100 915-918, 1986

7 Feinleib M, Zelen M Some pitfalls in the evaluation of screening programs *Arch Envir Health* 19 412 - 415, 1969

8 Fontana RS, Sanderson DR, Taylor WF et al Early lung cancer detection results of the initial (prevalence) radiologic and cytologic screening in the Mayo Clinic Study *Am Rev Respir Dis* 130 561 - 565, 1984


10. Markov Modelling in Clinical Trial Design

Illustrated by a Trial to Assess the Value of Bronchoscopy in the Early Detection of Lung Cancer Following Laryngeal Cancer

10.1 Abstract

This chapter studies the use of the Markov model in the design of a randomized clinical trial (RCT) to estimate the efficacy of a new medical treatment compared with the standard treatment. The Markov model predicts a survival curve of patients for either treatment group, using time dependent probabilities derived from (a) disease related mortality rates taken from a retrospective study, (b) age-specific mortality rates and (c) assumptions about the benefit of the new treatment. The Markov model can effectively be used to test the sensitivity of the efficacy of the new treatment for variation of the parameters used in the model.

The efficacy of the new treatment is expressed both as the improvement in survival predicted by the Markov model, and as the sample size of an RCT to detect the improvement. The improvement itself is an indicator of the clinical importance of the new treatment, the sample size is an indicator of the feasibility of the trial. Sample sizes are based on comparison of the survival curves predicted by the Markov model, rather than on comparison of single survival rates.

The method is illustrated by an RCT to assess the value of bronchoscopy in the early detection of lung cancer in patients with laryngeal cancer.

10.2 Introduction

PRINCIPLES OF RANDOMIZED CLINICAL TRIALS

A randomized clinical trial (RCT) is a clinical experiment to assess the efficacy of a new diagnostic or therapeutic treatment by randomization i.e. by random assignment of either the new treatment or the old 'standard' treatment to consecutive patients. The treatment groups of patients, also denoted as 'treatment arms' or 'trial arms', are compared afterwards. The treatment that, to a statistically significant extent, yields the best results, for instance the best survival rate, is said to

---

1 L J A Stalpers, J H M Zwetsloot-Schonk, A L M Verbeek and W A J van Daal For publication in Medical Decision Making
be the superior treatment. If the study yields no significant differences, the treatments are said to be equally effective. The latter statement is only valid if the study is large enough not to miss a clinically important difference.

A major pitfall of an RCT concerns the discrepancy between the benefit of a treatment one wants to detect, the benefit one can expect and the subsequent sample size to detect it [Freiman et al., 1978, Haybittle 1988, Pocock SJ 1978]. Therefore, in the designing phase of an RCT one should define which benefit is considered clinically important, estimate the potential benefit and estimate the feasibility of the trial considering the number of patients required and available to detect the benefit.

Two features may encumber the estimation of the efficacy of a new treatment by an RCT.

(a) **Effect dilution** It is generally unknown who amongst all treated patients will eventually profit from the new treatment. For instance, adjuvant chemotherapy in breast cancer is aimed at elimination of subclinical tumour dissemination, so-called micrometastasis in women with locoregional tumour spread. Since we do not know in advance who actually has micrometastasis, the effect of adjuvant chemotherapy has to be measured in all women with locoregional disease. A potential benefit in a small number of patients may then become insignificant for the larger group of all treated patients.

(b) **Effect timing** It can be difficult to estimate how long we should observe our patients in order to detect the potential benefit. This latter feature is encumbered by the fact that the prognosis of a patient changes. In general, the mortality risk from cancer decreases from the onset of treatment, but mortality from other causes increases.

THE MARKOV MODEL FOR SIMULATING RCT’s

The Markov model of prognosis [Beck & Pauker, 1983] can be used to simulate an RCT and to estimate the outcome of either arm. The Markov model has some attractive features compared with a simple probability tree analysis (comparable to a decision tree analysis). Markov modelling can be used to represent decision problems in a dynamic way. This is especially useful in structuring the course of a disease with time varying probabilities. Hence, rather than estimating simple endpoints such as five-year survival rates, a Markov model can more accurately generate whole survival curves for comparison. Furthermore, the Markov model is a
powerful tool to test the strength of conclusions by simulating the course of disease under various conditions. Hence, it can effectively deal with several presumptions about the benefit of the new treatment compared with the standard treatment.

For a certain level of significance and power of study and for a given (or hypothesized) difference between the survival curves produced by the Markov model, the sample size of an RCT can be calculated.

EXAMPLE: BRONCHOSCOPY IN PATIENTS WITH LARYNGEAL CANCER

The use of a Markov model of prognosis in a clinical trial design is demonstrated by a European RCT investigating the efficacy of bronchoscopy for early detection of lung cancer in patients with cancer of the head and neck [Pastorino et al., 1987]. This trial is presented here for illustrative purposes only. In the European trial a standard follow-up protocol which includes yearly chest-radiography is compared with a new protocol which includes frequent bronchoscopy to detect secondary lung cancer in a resectable stage. It is assumed that by frequent bronchoscopy the mortality from lung cancer will be reduced which eventually results in a better survival in patients with laryngeal carcinoma. The Markov model provides an estimate of the reduction of the mortality rate of patients diagnosed with laryngeal cancer due to bronchoscopy.

The probabilities used in the model are derived from (1) the incidence rate of lung cancer, (2) the disease-specific mortality rate from laryngeal cancer, (3) the mortality rate from secondary lung cancer. These rates were estimated by data from an own retrospective study of patients with laryngeal cancer [Engelen et al., 1990]. The model further includes (4) assumptions about the effect of bronchoscopy on the survival from lung cancer and (5) mortality rates from death of intercurrent diseases. Assumptions about the benefit of bronchoscopy, expressed as reduced lung cancer mortality rates, are based on some population-based studies drawn from the literature. Mortality rates from intercurrent diseases are based on age-specific death rates for the Dutch population [CBS, 1987].

AIM OF THE STUDY

The aim of this study is to show the use of the Markov model to estimate the efficacy of a new treatment by an RCT. The efficacy is expressed as the improvement in survival predicted by the Markov model, and as the sample size of an RCT to detect the improvement. The sample size is expressed as the yearly number of patients.
required for randomization. The improvement itself is an indicator of the clinical importance of the new treatment, the sample size is an indicator of the feasibility of the trial. The range of the inaccuracy of the model is explored in a sensitivity analysis on the parameters used in the model.

10.3. Data and Methods

SOME FACTS AND FIGURES ON LARYNGEAL CANCER AND LUNG CANCER

In Western countries laryngeal cancer is diagnosed in 114 per 1,000,000 in men and 6 per 1,000,000 in women per year [Waterhouse et al, 1982]. It is predominantly a disease of the sixth decade [Zagars & Norante, 1982]. Laryngeal carcinoma gives relatively early rise to complaints and cure can be relatively easily achieved both by radiotherapy and surgery [Stalpers et al, 1987a, Stalpers et al, 1987b, Stalpers et al, 1988].

A considerable number of patients cured from laryngeal cancer will develop lung cancer, probably because of smoking as a common etiological factor [Wynder et al, 1977]. Lung cancer following laryngeal cancer is seen more frequently in male patients than in female ones and more frequently in patients with supraglottic cancer than with glottic cancer [Wagenfeld et al, 1980, Wagenfeld et al, 1981]. Since the first study on second primary malignancies by Warren and Gates in 1932, increased occurrences of lung cancer following laryngeal carcinoma have frequently been described in the literature [Warren & Gates, 1932]. The occurrence has been described to range from 2.0% to 12.0%, depending on the duration of follow-up in the different studies and the definition applied to differentiate a primary malignancy in the lung from a lung metastasis [Martin et al, 1979, Hordijk et al, 1985, Cahan, 1977].

The overall five-year survival rate of patients with lung cancer ranges from 7% to 11% [Ries et al, 1983, Nou, 1984]. Prognosis is related with histology and with disease stage. For the prognostically favourable clinical stage I squamous cell carcinoma, the five-year survival may be 40% [AJC, 1979]. Most patients die within two years, after a five-year survival a patient can be considered cured.

Regular lung-radiography is performed in the follow-up of patients with laryngeal cancer in most oncology centres. Chest-radiography alone fails to detect lung cancer in an early stage in patients with cancer of the aerodigestive tract [Gluckman & Grissman, 1983]. These findings agree with the disappointing
experiences with population screening for lung cancer [Bailar, 1984; Fontana et al., 1984].

Bronchoscopy has been proposed to enhance the early detection of lung cancer. Bronchoscopy is an invasive diagnostic procedure which poses a substantial burden to the patient but which, in experienced hands, is said to be safe [Leipzig et al., 1985]. The test-characteristics of bronchoscopy are good when used to confirm radiographically suspect malignancies [Cortese et al., 1983; Mathisen et al., 1984]. However, it is not clear whether bronchoscopy will be useful in patients without manifest lung cancer [Leipzig et al., 1985].

RETROSPECTIVE STUDY: INCIDENCE RATES AND MORTALITY RATES
We used data from an own retrospective study to obtain actuarial estimates for (1) the incidence rate of lung cancer, (2) the disease specific mortality rate from laryngeal cancer and (3) the mortality rate from secondary lung cancer [Cutler & Ederer, 1958; Engelen et al., 1990].

From January 1979 to January 1986, 556 patients with a newly diagnosed squamous cell carcinoma of the larynx have been treated and followed at the Nijmegen University Hospital, The Netherlands. The follow-up study closed in June 1986, resulting in a minimum of 6 months and a maximum of 90 months follow-up (mean=36 months). Age ranged from 32 to 96 years with a mean age of 61 years. The male:female ratio was 10.5. Tumours were classified according to the TNM classification of tumours [AJC, 1978; UICC, 1978] and were localized in the glottic region in 317 patients (57%), in the supraglottic region in 234 patients (42%) and in the subglottic region in 5 patients (1%).

In general, primary treatment consisted of radiotherapy (60-74 Gy in 5 weeks) for T1 and T2 glottic and for T2 supraglottic carcinomas. Surgery was performed as primary treatment in all other patients and as salvage treatment in patients with a loco-regional recurrence. Chemotherapy was reserved for palliative treatment. The follow-up protocol was similar to the standard protocol in the projected RCT, including regular physical examination of the head and neck and yearly radiography of the lungs.

THE MARKOV MODEL
A Markov model with time dependent transition probabilities was used to calculate the expected survival curve of patients with laryngeal cancer for each trial arm. The
Markov model includes annual estimates of the incidence rate of lung cancer following laryngeal cancer and death rates from laryngeal cancer, from secondary lung cancer with and without bronchoscopy and from intercurrent diseases. In order to estimate the maximum possible benefit from bronchoscopy, it was assumed that all patients with lung cancer in the bronchoscopy group were detected in the most favourable clinical stage I. The yearly required number of patients for randomization is calculated based on the expected reduction of the mortality rate after five years follow-up due to bronchoscopy.

Structure of the Markov Model

Figure 10.1 gives the basic structure of the Markov model. The model is structurally identical for the standard protocol as well as for the bronchoscopy protocol. Of course, the transition probabilities are different for both arms.

Following randomization, all patients start in a state called “WELL”. In the first year, a patient either dies of intercurrent disease or of laryngeal cancer ending in the state called “DEAD” or stays alive. Some of these patients get lung cancer and
either die of lung cancer before the end of the year or survive. These patients will start the next year in a state called "LUNG CANCER." The remaining patients stay well for the rest of the year, remaining in the state WELL again in the next year. A patient starting with "LUNG CANCER" at year=1, may either die of laryngeal cancer, intercurrent disease or lung cancer or may survive. As the risk of dying of lung cancer depends on the time elapsed since lung cancer was diagnosed, and not on the time of randomization, we defined separate lung cancer states for the first five years post lung cancer to enable modelling of a decreasing lung cancer mortality rate (not represented in figure 10.1).

The events in the first four years following lung cancer are structurally identical. However, the lung cancer mortality rate diminishes each year. After five years following lung cancer detection, patients can be considered cured of lung cancer. Therefore, we assumed that these patients are "WELL" five years after the diagnosis of lung cancer.

Summarizing, we framed the problem in a seven-state Markov model, including four separate lung cancer states:

1. WELL: Alive, without lung cancer
2-6. LUNG CANCER: Alive, with lung cancer
7. DEAD: Dead, either of lung cancer, laryngeal cancer or of other causes

The survival at any year following laryngeal cancer is the proportion of patients alive, both alive in the WELL state and alive in one of the LUNG CANCER states. The mortality rate after T years follow-up is the proportion fatal Markov transitions (=deaths) of all transitions that could have ended fatally (=patient numbers at risk).

**Assignment of Probabilities**

The transition probabilities are not constant for all years. With the years the risk of dying from laryngeal cancer, the risk of getting lung cancer and the risk of dying from lung cancer decreases, while the risk of dying from other "age-related" causes increases. So the transition probabilities in our model are time-dependent.
estimate the transition probabilities we used data from the retrospective study and from the literature. Table 10.1 gives a summary of the available data.

Table 10.1 • Summary of available data

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>Five-year cumulative incidence of secondary lung cancer in patients with laryngeal cancer*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• All patients</td>
<td>17%</td>
</tr>
<tr>
<td></td>
<td>- female</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>- male</td>
<td>18%</td>
</tr>
<tr>
<td></td>
<td>• Supraglottic</td>
<td>22%</td>
</tr>
<tr>
<td></td>
<td>- male</td>
<td>24%</td>
</tr>
<tr>
<td></td>
<td>• Glottic</td>
<td>13%</td>
</tr>
<tr>
<td></td>
<td>- male</td>
<td>14%</td>
</tr>
<tr>
<td>(2)</td>
<td>Laryngeal cancer five-year survival rates, corrected for secondary lung cancer incidence and age-specific death rate (see text)*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• All patients</td>
<td>87%</td>
</tr>
<tr>
<td></td>
<td>- male</td>
<td>88%</td>
</tr>
<tr>
<td></td>
<td>• Supraglottic</td>
<td>87%</td>
</tr>
<tr>
<td></td>
<td>- male</td>
<td>88%</td>
</tr>
<tr>
<td></td>
<td>• Glottic</td>
<td>88%</td>
</tr>
<tr>
<td></td>
<td>- male</td>
<td>89%</td>
</tr>
</tbody>
</table>

Five-year survival from lung cancer

(3) "Standard Protocol" 21%
(4) "Bronchoscopy Protocol" 40%#

(5) Mean age at randomization = 61 years

* Based on the retrospective study [Engelen et al., 1990]
# Based on the literature [AJC, 1979]
(1) Lung cancer incidence rate (LIR)
In the retrospective study, 69 out of 556 patients with laryngeal cancer (12.4%) developed secondary lung cancer. After actuarial correction for the duration of follow-up, the cumulative incidence was 17% after five-year [Engelen et al., 1990]. In order to enable sensitivity analysis of the yearly incidence rate of lung cancer we approximated the incidence rate by a simple linear function:

\[
\text{lung cancer incidence rate}_t = \text{LIR}_t = a - b \cdot t
\]

where \(a\) denotes the initial incidence rate and \(b\) the regression coefficient, with the restriction that if \(\text{LIR} \leq 0\) then \(\text{LIR} = 0\). The cumulative incidence \(\text{CI}\) of secondary lung cancer at time \(i \geq 1\) is given in (2):

\[
\text{CI}_i = \text{CI}_0 + \sum_{t=1}^{i} \{\text{LIR}_t \cdot (1.0 - \text{Cl}_{t-1})\}
\]

Using least squares statistics, best curve fit to our data (correlation = 0.88) was obtained for:

\[
\text{LIR}_t = 0.063 - 0.009t
\]

Following stratification by primary site of the laryngeal tumour, the cumulative incidence is lower after glottic than after supraglottic cancer [Engelen et al., 1990]. The best fitting function for patients with glottic carcinoma was:

\[
\text{LIR}_t = 0.0541 - 0.0089t
\]

and for patients with supraglottic carcinoma:

\[
\text{LIR}_t = 0.0814 - 0.0104t
\]

In the retrospective study, none of the 54 women developed lung cancer. Two women with verified lung metastasis from breast cancer were considered as not having lung cancer. Therefore we modelled the lung cancer incidences also for male patients only (see table 10.2).

(2) Laryngeal cancer mortality rate
The “cause-specific” mortality rate (CMR) of laryngeal cancer was calculated by correcting the crude mortality rate, observed in the 556 patients, for death of lung
cancer and for death of age and sex-related causes (ASR). The mean age of patients at diagnosis laryngeal cancer was 61 years. Correction for the ASR was performed by subtracting the age-specific mortality rate for the Dutch population aged 61 [CBS, 1987] at diagnosis laryngeal cancer from the crude mortality rate. We modelled the cause-specific mortality rate from laryngeal cancer by a simple linear function:

$$CMR_t = c - d \cdot t$$

were \(c\) is the initial mortality rate. We made the restriction that \(CMR_t \leq 0\) than \(CMR_t = 0\). Best fit was obtained for \(c = 0.064\) and \(d = 0.0108\) (correlation = 0.99) ²

### (3) Lung cancer mortality rate

Figure 10.2 shows the survival curve of 69 patients with lung cancer following laryngeal cancer. The five-year survival is 21%. Corresponding with the literature on lung cancer survival, the survival curve remains at the same level after five years. The survival curve could best be modelled if we presumed a decreasing exponential lung cancer mortality rate:

$$\text{lung cancer mortality rate} = e^{-\lambda t}$$

The survival \(S\) at any time \(t\) following lung cancer can then be calculated with formula (5):

$$S(t) = S(0) - \sum_{t=1}^{i} \left[ e^{-\lambda t} \times S(t-1) \right] \quad (i \geq 1; S(0)=1)$$

The \(\lambda\) cannot simply be calculated, but has to be approximated based on \(S(t)\).

² The data on laryngeal cancer mortality in table 10.1 show an unexpected feature: The cause-specific survival from supraglottic cancer, i.e. corrected for death from age and sex-related causes and lung cancer-, showed to be only marginally lower than the survival from glottic cancer. However, the overall survival in patients with supraglottic cancer is significantly worse than in patients with glottic cancer, which agrees with the literature [Zagars & Norante, 1983]. The disagreement between cause-specific and overall survivals could not be explained by a different distribution of age, sex or chance. It could be explained by the substantially higher incidence of (and subsequent death from) lung cancer in patients with supraglottic cancer.
Based on the five-year survival of 0.21 and using equation (4) a decreasing exponential mortality rate for patients with lung cancer following laryngeal cancer was approximated with $\lambda = 0.58$ in formula (5) The lung cancer mortality rate in the first year then will be $e^{-0.58 \times 1} = 0.56$ and in the second year $e^{-0.58 \times 2} = 0.31$ etc.

In order to estimate the optimal effect of bronchoscopy, we presumed that by bronchoscopy all patients with lung cancer have a squamous cell carcinoma detected in the most favourable clinical stage = I [AJC, 1978, 1979, UICC, 1986] Assuming a 40% five-year survival of these patients, $\lambda$ is 0.84 The lung cancer mortality in the first year then will be $e^{-0.84 \times 1} = 0.43$ and in the second year $e^{-0.84 \times 2} = 0.19$ etc

We did not correct the lung cancer mortality rates for age-specific death rate or death of laryngeal cancer, as these mortality rates were negligible compared to the lung cancer mortality rate for the first five years. To correct for state transitions during the year, we made half-time corrections of the transition probabilities (see appendix I, §10.6)

Table 10.2 gives a summary of the baseline values for the probabilities used in the analysis.

Figure 10.2 *Actuarial survival from lung cancer in 69 patients following laryngeal carcinoma (broken line, patient numbers at risk are indicated) and the curve based on eq (5) used in the Markov model (solid line)*
Table 10.2 • Summary of the baseline parameters used in the analysis

<table>
<thead>
<tr>
<th>Probability</th>
<th>Assumed time-dependent model</th>
<th>Correlation with original data</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Lung cancer incidence rate</td>
<td>$a - b \cdot t$ per year</td>
<td></td>
</tr>
<tr>
<td>• All patients</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- male</td>
<td>$0.063$</td>
<td>$0.0090$</td>
</tr>
<tr>
<td></td>
<td>$0.0711$</td>
<td>$0.0103$</td>
</tr>
<tr>
<td>• Supraglottic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- male</td>
<td>$0.0814$</td>
<td>$0.0104$</td>
</tr>
<tr>
<td></td>
<td>$0.0917$</td>
<td>$0.0120$</td>
</tr>
<tr>
<td>• Glottic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- male</td>
<td>$0.0542$</td>
<td>$0.0089$</td>
</tr>
<tr>
<td></td>
<td>$0.0574$</td>
<td>$0.0094$</td>
</tr>
<tr>
<td>(2) Laryngeal cancer mortality rate</td>
<td>$c - d \cdot t$ per year</td>
<td></td>
</tr>
<tr>
<td>• All patients</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- male</td>
<td>$0.0635$</td>
<td>$0.0108$</td>
</tr>
<tr>
<td></td>
<td>$0.0629$</td>
<td>$0.0121$</td>
</tr>
<tr>
<td>• Supraglottic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- male</td>
<td>$0.0655$</td>
<td>$0.0134$</td>
</tr>
<tr>
<td></td>
<td>$0.0765$</td>
<td>$0.0161$</td>
</tr>
<tr>
<td>• Glottic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- male</td>
<td>$0.0604$</td>
<td>$0.0109$</td>
</tr>
<tr>
<td></td>
<td>$0.0624$</td>
<td>$0.0116$</td>
</tr>
<tr>
<td>Lung cancer mortality rate</td>
<td>$e^{-\lambda t}$ per year</td>
<td></td>
</tr>
<tr>
<td>(3) • &quot;Standard Protocol&quot;</td>
<td>$\lambda = 0.58$</td>
<td></td>
</tr>
<tr>
<td>(4) • &quot;Bronchoscopy Protocol&quot;</td>
<td>$\lambda = 0.84$</td>
<td></td>
</tr>
<tr>
<td>(5) Death of intercurrent diseases</td>
<td>Increasing age-specific mortality rate</td>
<td>Age at randomization</td>
</tr>
<tr>
<td></td>
<td>61 year</td>
<td></td>
</tr>
</tbody>
</table>
Validating the Markov Model

Since mathematical modelling of all the transition probabilities might have lead to a deviation of the Markov model from the original data upon which it was based, we compared (1) the actuarial survival curve with (2) the survival curve as predicted by the Markov model for patients in the standard protocol. As shown in figure 10.3 and table 10.3, the differences are marginal.

CALCULATIONS OF REQUIRED PATIENT NUMBERS

The calculation of the number of patients yearly required for randomization, is based on the expected differences in average mortality rates after five years recruitment and follow-up. The numbers are calculated for a bi-armed randomized clinical trial comparing the standard protocol with the bronchoscopy protocol for a study with a two-tailed significance $\alpha = 0.05$ and a one-tailed power $1-\beta = 0.90$ [Bulpitt, 1983]. A significance of 0.05 means that there is only 5% chance that the trial will demonstrate a difference between the two protocols where there is in fact no difference. A power of 0.90 means that there is a probability of 0.9 of demonstrating a benefit of bronchoscopy if there really is a benefit over the standard follow-up.

Halperin et al. (1968) described a method to calculate the required patient numbers in an RCT in order to detect the expected difference between two simple proportions with a certain significance and study power.
In follow-up studies such as most cancer clinical trials, one usually compares whole survival curves, rather than single endpoints such as the five-year survival rate. The survival curves produced by the Markov model can be treated as actuarial curves with one-year intervals. In analogy with the logrank test to compare survival curves [Mantel, 1966], we can define an average annual mortality rate for the duration of follow-up. Using the Markov probability distribution of states $p_t(\text{ALIVE, DEAD})$ for each year, we used an adjusted version of the Halperin method to calculate the annual number of patients for randomization based on the difference in average mortality rate for a certain duration of follow-up. The method is described in appendix II (§10.7).

Table 10.3 • Validation of the Markov model: Actuarial five-year survival and predicted five-year survival for patients with laryngeal carcinoma by sex and localization of the laryngeal tumour

<table>
<thead>
<tr>
<th>Patient group</th>
<th>Standard protocol</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Actuarial 5-year survival (%)</td>
<td>Predicted 5-year survival (%)</td>
</tr>
<tr>
<td>All</td>
<td>68.5</td>
<td>67.74</td>
</tr>
<tr>
<td>• male</td>
<td>68.0</td>
<td>68.32</td>
</tr>
<tr>
<td>Supraglottic</td>
<td>66.4</td>
<td>66.53</td>
</tr>
<tr>
<td>• male</td>
<td>65.4</td>
<td>64.16</td>
</tr>
<tr>
<td>Glottic</td>
<td>71.32</td>
<td>70.79</td>
</tr>
<tr>
<td>• male</td>
<td>71.33</td>
<td>70.41</td>
</tr>
</tbody>
</table>

10.4 Results

BASELINE ANALYSIS
As shown in figure 10.3, a small improvement of the five-year survival of patients with laryngeal cancer thanks to regular bronchoscopy is predicted by the Markov model. The exact improvement 2.1% from 67.7% to 69.8%. After five years follow-up, the average annual mortality rate was reduced from 8.21% to 7.72%. It can be
calculated that at least 9,550 patients are required for randomization every year to detect the predicted improvement (for calculation: See table 10.6 in Appendix II, p.136).

In the retrospective study we found that in patients with cancer of the glottic and supraglottic larynx after five years 13% and 22% respectively had developed lung cancer. The number of patients required for randomization is calculated for each patient group. As may be read from table 10.4, the predicted improvement in five-year survival ranges from 1.68% to 3.05%. The reduction in average annual mortality rate ranges from 0.40% to 0.75%. As we found no lung cancer in women in the retrospective study, no profit is to be expected from bronchoscopy for women.

Table 10.4 · Baseline analysis: Predicted improvement in five-year survival rate, the reduction in average annual mortality rate after five years follow-up and the patient numbers required in a bi-armed randomized clinical trial to detect this improvement (see text).

<table>
<thead>
<tr>
<th>Patient group</th>
<th>Predicted improvement in 5-year survival (%)</th>
<th>Predicted mortality reduction (%)</th>
<th>Number of patients required</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>2.08</td>
<td>0.49</td>
<td>9550</td>
</tr>
<tr>
<td>• male</td>
<td>2.33</td>
<td>0.55</td>
<td>7509</td>
</tr>
<tr>
<td>Supraglottic</td>
<td>2.81</td>
<td>0.67</td>
<td>5356</td>
</tr>
<tr>
<td>• male</td>
<td>3.05</td>
<td>0.75</td>
<td>4715</td>
</tr>
<tr>
<td>Glottic</td>
<td>1.68</td>
<td>0.40</td>
<td>13013</td>
</tr>
<tr>
<td>• male</td>
<td>1.77</td>
<td>0.43</td>
<td>11496</td>
</tr>
</tbody>
</table>

SENSITIVITY ANALYSIS
The number of patients required in an RCT is based on the reduction in average annual mortality rate in a study with a follow-up of five years, a power of 0.90 and a significance of 0.05. These may be relatively stringent criteria. We therefore performed sensitivity analysis on (a) the power of the study trial and (b) the duration of the follow-up.
The retrospective study was based on a patient population from only one specific university hospital. To account for variations in populations from the different hospitals involved in the multicentre trial, we performed sensitivity analysis on (c) lung cancer survival and (d) age at randomization.

(a) Power of the study
Table 10.5 shows a sensitivity analysis on the power of the study for the number of patients required in the RCT. For instance, if the study would have a power 0.80, which is said to be acceptable in many studies, the number of patients required would be 6900 to detect a 0.5% mortality reduction from 8.21% to 7.71% per year.

(b) Duration of follow-up
In the baseline analysis it was assumed that the study duration was five years, both for the follow-up and randomization. However, the benefit of early detection of lung cancer may become more prominent if we would take a longer duration of follow-up. The Markov model easily allows us to predict survival improvements for shorter or longer follow-up periods. Table 10.6 in appendix II (§10.7) shows the predicted survival improvement for a follow-up ranging from 0 to 20 years. The mortality reduction increases up to 7.92% - 7.41% = 0.51% after 6 years follow-up. After this...
period, the mortality from age-specific diseases becomes larger than the mortality from lung cancer, and the reductions will gradually decrease. (This finding agrees with the trivial knowledge about human existence that in the long run, everyone will die.) From table 10.6 we can read that even for a follow-up of 10 years, at least 2573 patients with laryngeal cancer are yearly required for randomization to detect the mortality reduction.

(c) Lung Cancer Survival

In the retrospective study, the five-year survival of patients with lung cancer following laryngeal cancer was 21%. If we presume worse five-year lung cancer survival rates in the standard protocol, more patients may profit from bronchoscopy. Hence, the number of patients required in a randomized trial will be lower, providing of course that by bronchoscopy all patients with lung cancer will be found having a clinical stage I carcinoma with a 40% five-year survival.

Figure 10.4 shows the predicted survival improvement from laryngeal carcinoma for five-year lung cancer survival ranging from 0% to 25%. From this figure we can read that in a study with a power of 90% the required patient numbers range from 2235 to 14980.

For the baseline analysis it was assumed that, following bronchoscopy, all patients will be detected as clinical stage I, which is highly optimistic. In stage II, the five-year survival of lung cancer will drop to almost 20% [AJC, 1979]. As far as we know, it is...
not yet well established in what stages lung cancer is detected by bronchoscopy. We therefore assume that the actual five-year survival ranges between 20% and 40%, depending on the distribution of stages found in the bronchoscopy group. For instance, if both stage I and stage II are equally represented, the five year survival of the bronchoscopy group will be the average = 30%. Figure 10.5 gives the sensitivity analysis of the required patient numbers for various survival rates for lung cancer in the bronchoscopy group. For a study power of 80% and a five-year survival of 30%, it can be read that 44070 patients are annually required.

Figure 10.5 • Sensitivity analysis of the required number of patients for various five-year survival rates from lung cancer in the bronchoscopy group ranging from 25 to 40% and for various study powers

(d) Age at Randomization
The mean age of the retrospective study group was 61 years. Because regular bronchoscopy is very inconvenient for older patients, they may be excluded from the study group. Hence the mean age will be lower in an actual study group.

Figure 10.6 shows the predicted survival improvement and required number of patients for varying mean ages of the study population at randomization. Since at a younger age mortality is less determined by age-related diseases and more by lung cancer mortality, a beneficial effect of bronchoscopy is higher at a younger age than at an older age. This is expressed in a larger predicted improvement and correspondingly lower numbers required in a trial.
10.5 Discussion

This study describes the use of Markov modelling in the design of a randomized clinical trial to forecast the benefit of a new treatment in terms of survival improvement and to estimate the number of patients required to detect the benefit.

The Markov model is a powerful tool to predict patient prognosis with time-dependent probabilities [Beck & Pauker, 1983, Ransohoff et al., 1983]. As in any model, one of the greatest limitations for its application lies in the reliability of the data and the simplifications of the real-world problem [Pauker & Kassirer, 1987].

We used data from an own retrospective study to estimate probabilities. This approach enables accurate modelling of time-dependent probabilities with actuarial data. Rather than imposing a certain “biologically sound” model on our data, we preferred to modelling by “best fit” procedures. This simple approach may conflict with some biological assumptions for which extra restrictions have to be made (such as that an incidence-rate may never become negative). Since an accumulation of simplifications may eventually lead to an increasing deviation from the “real world”, we compared the survival curve predicted by the model with the actuarial survival curve of the original data-set. Especially compared with a more simple “classical” probability tree using five years' results, the Markov model produces accurate estimates of the actuarial survival estimates. It is evident, that external validation of the model is only possible by a check with another data-set.

The estimation of the highest benefit of the new treatment, namely by reduction of the lung cancer mortality to the lowest realistic rate due to regular
bronchoscopy-, was based on data taken from the literature, namely on the lung cancer survival rate in the prognostically most favourable clinical stage [AJC, 1979]. A sensitivity analysis was performed to analyse the efficacy of the new treatment under less and unrealistically more favourable conditions.

Since the annually required number of patients is one of the most important factors determining the feasibility of an RCT, we used this number as an index for comparison. Rather than comparing single end-points such as the five-year survival rate, we compared the average mortality rates drawn from whole survival curves as predicted by the Markov model. The comparison of the average mortality rates is the essence of the logrank test as proposed by Mantel (1966). This approach does not only yield more powerful and accurate information of survival, but can also be used to analyse the efficacy of the new treatment for varying durations of follow-up.

The feasibility of an RCT does not merely depend on the number of patients required, but also on the availability of patients. In the European trial on which this study is based, it was suggested that within five years about 2,000 patients would be required for randomization [Pastorino et al., 1987]. Under the baseline assumptions given in table 10.3, this number will be insufficient to demonstrate a five-year survival improvement. Although the reduction in lung cancer mortality rate may be huge, the effect is so "diluted" by the whole group of patients with laryngeal cancer, that the eventual mortality rate is only marginally reduced due to bronchoscopy. To detect such small reductions as estimated by the Markov model with an acceptable power, it requires a long follow-up duration in a study with a (too) large sample size. A sensitivity analysis on the lung cancer survival rates or the average age at randomization shows only marginal variations in predicted improvements.

A randomized clinical trial (RCT) is widely advocated to assess the value of a new diagnostic or therapeutic treatment [Grage et al., 1982]. The method described in this study may provide some necessary information in the design of an RCT. The decision whether or not to embark on an RCT depends on whether the participating institutes judge it worthwhile to detect a probably small benefit of a new treatment, to invest the effort to recruit the required number of patients and to invest the time to follow these patients for a sufficient length of time.
Acknowledgment

The patients described in this study were treated and followed by the Head and Neck Oncology Group and the Lung Cancer Group of the St Radboud University Hospital, Nijmegen, The Netherlands, wherein participating the:
Departments of Otorhinolaryngology (head: Prof. Dr. P. van den Broek),
Maxillofacial Surgery (head: Prof. Dr. H.P.M. Freihofer),
Medical Oncology (head: Prof. Dr. D.J.Th. Wagener),
Diagnostic Radiology (head: Prof. Dr. J.H.J. Ruys),
Pulmonology (head: Prof. Dr. C.L.A. van Herwaarden) and
Radiotherapy (head: Prof. Dr. W.A.J. van Daal).

We are especially obliged to A. Engelen MB, for his help in collecting patient data and to Mrs. C.W. McKell for her linguistic help. We kindly acknowledge Dr. O. Dalesio from the Netherlands Cancer Institute, Amsterdam, and the two anonymous referees for their critical comments.

10.6 Appendix I: Half-time correction

In a Markov model frequencies of transition states are calculated at the end of each cycle. Transitions may occur over the year. Patients may develop lung cancer in the first year and may die within that same year. To correct for this double transition in the same year, we assume that on the average the transition from WELL to LUNG CANCER occurs in the middle of the year and that a patient is a half year at risk of dying of lung cancer. For patients in the standard protocol, the risk of dying of lung cancer within the first year if lung cancer has occurred = 0.5e^{-\lambda x 0.5}. Then, with respect to the follow-up of lung cancer, the lung cancer mortality rate in the second year equals e^{-\lambda x 1.5}.

Consequently, the risk of dying of lung cancer in the i-th year following the year wherein lung cancer was detected = e^{-\lambda (i-0.5)}. Correspondingly, the lung cancer survival curve as implemented in the Markov model will be an adjusted version of formula (5):

\[ S_{(i)} = S_{(0.5)} \times (1.0 - 0.5e^{-\lambda x 0.5}) - \sum_{t=2}^{i} \left[ e^{-\lambda (t-0.5)} \times S_{(t-1)} \right] \] (6)
where \( S(0.5) \) = the initial proportion of patients developing lung cancer within the first year and \( i \geq 1 \)

10.7 Appendix II Sample size

A method to calculate the required number of patients in a bi-armed randomized trial is described by Halperin et al. (1968). For a certain significance \( \alpha \) and power \( 1-\beta \), the number \( N \) of patients in each trial arm is based on the difference between simple proportions \( p_1 \) and \( p_2 \) (e.g., the 5-year survival) in each trial arm

\[
N = \left( \frac{Z_\alpha \sqrt{2P(1-P)}}{\rho_0} + Z_\beta \sqrt{2p_1(1-p_1) + p_2(1-p_2)} \right)^2
\]

(1)

\( Z_\alpha \) = Normal deviate for the two-tailed significance \( \alpha \) of 0.05 = 1.96
\( Z_\beta \) = Normal deviate for the one-tailed 1-power \( \beta \) of 0.10 = 1.28
\( p_1 = \) Proportion in study arm 1
\( p_2 = \) Proportion in study arm 2
\( P = 0.5(p_1 + p_2) \)

The Halperin method is a simple equation to give a rough estimate of the required numbers. However, it disregards two important aspects of most cancer clinical trials:

1. In follow-up studies one usually compares whole survival curves, rather than single proportions. In general, less patients are required to detect differences by comparing the survival curves than by single proportions. In section 1 we give a method to calculate the required patient numbers for differences in survival curves predicted by the Markov model, starting with \( N \) patients in each arm.

2. Although a clinical trial is usually analysed as a cohort study after a certain time of follow-up, the eventual number of patients at risk depends both on the number of patients for randomization each period (usually a year) and on the time during which new patients can be rendered into the study. The longer the duration of randomization and follow-up, the less patients are required for randomization each period. In section 2 we give a method to calculate the required number of new patients for randomization per period based on the Markov probability distribution after a certain period of randomization and follow-up.
Section 1: Required numbers in a simple cohort study

The Markov model, as described in this paper, predicts the actuarial survival curve of a simple cohort starting with \( N \) patients. Analogously to the logrank statistic commonly used for comparing actuarial survival curves [Mantel, 1966], we can calculate and compare the average mortality rates \( p \) of each survival curve for a follow-up of \( T \) years:

\[
P = \frac{n_T}{N_T} \tag{A.2}
\]

where \( n_T \) denotes the number of transitions (=deaths) and \( N_T \) the number of potential transitions (=total number of patients at risk) during period \( T \). Since more patients are at risk during the early years of follow-up than during later years, the mortality rates of those early years have a correspondingly greater weight in the calculation of the average mortality rate. The number of patients at risk for year \( t \) is:

\[
N_t = \hat{p}_{t-1}(\text{ALIVE}) \cdot N \tag{A.3}
\]

where \( \hat{p}_{t-1}(\text{ALIVE}) \) denotes the Markov probability of being ALIVE at time = \( t-1 \). For follow-up time \( T \), the total number of patients at risk is:

\[
N_T = \sum_{t=1}^{T} \hat{p}_{t-1}(\text{ALIVE}) \cdot N \tag{A.4}
\]

The number of deaths \( n_t \) in year = \( t \) is:

\[
n_t = p_t(\text{ALIVE} \rightarrow \text{DEAD}) \cdot N_t \tag{A.5}
\]

where \( p_t(\text{ALIVE} \rightarrow \text{DEAD}) \) denotes the transition probability (= mortality risk) of year = \( t \). The total number of deaths during period \( T \) is:

\[
n_T = \sum_{t=1}^{T} p_t(\text{ALIVE} \rightarrow \text{DEAD}) \cdot N \tag{A.6a}
\]

\[
= \hat{p}_T(\text{DEAD}) \cdot N \tag{A.6b}
\]
where $\hat{p}_T(\text{DEAD})$ denotes the Markov probability of being dead after $T$ years.

Substituting equations (A.4) and (A.6) in equation (A.2) yields the average mortality rate $p$ of a Markov survival curve:

$$p = \frac{n_T}{N_T} = \frac{\hat{p}_T(\text{DEAD})}{\sum_{t=1}^{T} \hat{p}_{t-1}(\text{ALIVE})} \quad (A.7)$$

Substituting the average mortality rates $p_1$ and $p_2$ drawn from the Markov survival curves of trial arm 1 and 2 in equation (A.1) now yields $N_T$. Substituting $N_T$ thus obtained from equation (A.1) in equation (A.4), yields the required number of patients $N$ in each arm in a trial lasting $T$ years:

$$N = \frac{n_T}{\sum_{t=1}^{T} \hat{p}_{t-1}(\text{ALIVE})} = \text{eq.(A.1)}$$

$$= \frac{1}{\sum_{t=1}^{T} \hat{p}_{t-1}(\text{ALIVE})} \left( \frac{Z_{\alpha}\sqrt{2P(1-P)}}{\sqrt{\sigma^2}} + \frac{Z_{\beta}\sqrt{[p_1(1-p_1) + p_2(1-p_2)]}}{\sqrt{(p_1 - p_2)^2 \sum_{t=1}^{T} \hat{p}_{t-1}(\text{ALIVE})}} \right)^2 \quad (A.8)$$

Section 2: Yearly required numbers in an RCT

In a randomized clinical trial new patients are yearly rendered into the study. Let us presume a trial with a follow-up of $T$ years during which every year $R$ new patients are rendered into each trial arm. In analogy to section 1, $R$ is calculated based on the average mortality rates $p_1$ and $p_2$ for arm 1 and 2 respectively, as predicted by the Markov model.

A trial in which every year new patients are entered during several years, can be seen as the cumulation of a cohort followed one year + another cohort followed two years + another followed three years etc.. Then, for a follow-up period of $T$ years, the total number of patients at risk can be calculated in analogy to equation (A.4):
\[ N_T = \sum_{t=1}^{T} (T-t+1) \cdot \hat{p}_{t-1}(\text{ALIVE}) \cdot R \]  

(A.9)

In analogy to equation (A.6b) the total number of death is:

\[ n_T = \sum_{t=1}^{T} \hat{p}_T(\text{DEAD}) \cdot R \]  

(A.10)

Substituting equations (A.9) and (A.10) in equation (A.2), yields the average mortality rate of a Markov survival curve with new patients for randomization each year during T years of follow-up:

\[ p = \frac{\sum_{t=1}^{T} \hat{p}_T(\text{DEAD})}{\sum_{t=1}^{T} (T-t+1) \cdot \hat{p}_{t-1}(\text{ALIVE})} \]  

(A.11)

Substituting the average mortality rates \( p_1 \) and \( p_2 \) for the Markov survival curves of trial arm 1 and 2 in equation (A.1) now yields \( N_T \). Substituting equation (A.1) for \( N_T \) in equation (A.9), yields the required number of patients \( R \) in each arm in a trial lasting \( T \) years:

\[ R = \frac{N_T}{\sum_{t=1}^{T} (T-t+1) \cdot \hat{p}_{t-1}(\text{ALIVE})} = \frac{\text{eq. (A.1)}}{\sum_{t=1}^{T} (T-t+1) \cdot \hat{p}_{t-1}(\text{ALIVE})} \]

(A.12b)

\[ = \left( \frac{Z_\alpha \sqrt{2P(1-P)}}{\frac{Z_\beta \sqrt{[p_1(1-p_1) + p_2(1-p_2)]}}{T} \cdot (p_1 - p_2)^2 \sum_{t=1}^{T} (T-t+1) \hat{p}_{t-1}(\text{ALIVE})} \right)^2 \]

Table 10.6 gives an example of the calculation of the number of patients yearly required in a bi-armed RCT for various durations of follow-up.
Table 10.6: Calculation of required patient numbers per year in a bi-armed randomized controlled trial for various duration of follow-up

<table>
<thead>
<tr>
<th>Duration of follow up (Years)</th>
<th>STANDARD</th>
<th>BRONCHOSCOPY</th>
<th>Mortality rate reduction</th>
<th>Required numbers at nsk</th>
<th>Yearly numbers required in arm 1 R1</th>
<th>Yearly numbers required in arm 2 R2</th>
<th>Yearly numbers required in an RCT R1 + R2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>p_t(ALIVE, DEAD)</td>
<td>1 0</td>
<td>0 0883</td>
<td>(1 00, 0 00)</td>
<td>(1 00, 0 00)</td>
<td>0 0023</td>
<td>315741</td>
</tr>
<tr>
<td>0</td>
<td>(912, 088)</td>
<td>0 088</td>
<td>1</td>
<td>0</td>
<td>0 0860</td>
<td>914, 086</td>
<td>0 086</td>
</tr>
<tr>
<td>1</td>
<td>(831, 169)</td>
<td>0 257</td>
<td>2 912</td>
<td>0 0884</td>
<td>(839, 161)</td>
<td>0 247</td>
<td>2 914</td>
</tr>
<tr>
<td>2</td>
<td>(765, 235)</td>
<td>0 492</td>
<td>5 654</td>
<td>0 0870</td>
<td>(778, 222)</td>
<td>0 469</td>
<td>5 667</td>
</tr>
<tr>
<td>3</td>
<td>(715, 285)</td>
<td>0 777</td>
<td>9 163</td>
<td>0 0848</td>
<td>(732, 268)</td>
<td>0 737</td>
<td>9 197</td>
</tr>
<tr>
<td>4</td>
<td>(677, 323)</td>
<td>1 100</td>
<td>13 386</td>
<td>0 0821</td>
<td>(698, 302)</td>
<td>1 039</td>
<td>13 461</td>
</tr>
<tr>
<td>5</td>
<td>(651, 349)</td>
<td>1 449</td>
<td>18 286</td>
<td>0 0792</td>
<td>(673, 327)</td>
<td>1 365</td>
<td>18 422</td>
</tr>
<tr>
<td>6</td>
<td>(627, 373)</td>
<td>1 822</td>
<td>23 837</td>
<td>0 0764</td>
<td>(651, 349)</td>
<td>1 715</td>
<td>24 056</td>
</tr>
<tr>
<td>7</td>
<td>(604, 396)</td>
<td>2 218</td>
<td>30 015</td>
<td>0 0739</td>
<td>(628, 372)</td>
<td>2 087</td>
<td>30 347</td>
</tr>
<tr>
<td>8</td>
<td>(581, 419)</td>
<td>2 637</td>
<td>36 797</td>
<td>0 0717</td>
<td>(604, 396)</td>
<td>2 483</td>
<td>37 257</td>
</tr>
<tr>
<td>9</td>
<td>(557, 443)</td>
<td>3 079</td>
<td>44 161</td>
<td>0 0697</td>
<td>(579, 421)</td>
<td>2 904</td>
<td>44 772</td>
</tr>
<tr>
<td>10</td>
<td>(415, 585)</td>
<td>5 707</td>
<td>88 802</td>
<td>0 0643</td>
<td>(431, 569)</td>
<td>5 438</td>
<td>90 490</td>
</tr>
<tr>
<td>15</td>
<td>(263, 737)</td>
<td>9 096</td>
<td>143 790</td>
<td>0 0633</td>
<td>(273, 727)</td>
<td>8 763</td>
<td>146 963</td>
</tr>
</tbody>
</table>

** = Σ(T-t+1)p_t-1(ALIVE)
Practical notes

If Decision Maker 6.0™ [Pauker et al., 1988] is used to analyse the Markov model, the average mortality rate $p$ can be calculated as follows:

1. Define DUAL bindings. The bindings for each state ALIVE is:
   - $m_{ulINIT} = \text{DUAL}(0, m \text{ CYCLE})$
   - $m_{ulNCR} = \text{DUAL}(0, m \text{ CYCMAX}-m \text{ CYCLE})$

   and for the DEAD state:
   - $m_{ulINIT} = \text{DUAL}(0,0)$
   - $m_{ulNCR} = \text{DUAL}(1,0)$

2. Follow-up time $T = m \text{ CYCMAX}$

3. A Cost-Based C-E Analysis yields:

   \[
   \sum_{t=1}^{T} (T-t+1) \cdot \hat{p}_{t-1}(\text{ALIVE}) = \text{COST}
   \]

   \[
   \sum_{t=1}^{T} \hat{p}_{t}(\text{DEAD}) = \text{EFF}
   \]

   \[
   p = \text{AVG C/E}
   \]

10.8 References

1. AJC-American Joint Committee for Cancer Staging and End Results Reporting. *Clinical Staging System for Carcinoma of the Larynx*. American Joint Committee (AJC) for Cancer Staging and End Results Reporting, Chicago 1978

2. AJC-American Joint Committee for Cancer Staging and End Results Reporting. *Staging of Lung Cancer, 1979 Manual of the AJC on cancer staging and end-results reporting task force on lung*. AJC, Chicago, 1979


11. Utility Assessment by Time Tradeoffs: A Delusive Deal

11.1 Abstract

Torrance's time tradeoff (TTO) technique has been promoted as a reliable and valid method to assess numerical values (utilities) for health-related quality of life issues. In a TTO-test the utility of a certain health state is complementary to the proportion that a subject is willing to tradeoff from the length of life in a worse health state. The TTO-technique is based on the major assumption that the proportion traded off is the same for any length of life in the worse state.

The assumption of proportionality was tested in 56 medical students and 9 physicians making tradeoffs between the quality of speech and the quantity of life in the choice between radiotherapy and surgery for cancer of the vocal cords. In order to live with a better quality of speech, subjects were generally willing to tradeoff a considerable bigger proportion of the length of life when faced with a long life expectancy than when faced with a short life expectancy. This study shows a severe violation of the proportionality assumption (p<0.01), giving rise to doubts about the validity of the TTO technique for the assessment of utilities.

11.2 Introduction

Since improvement of the quality of life forms an important goal of clinical and community health care, the assessment of numerical values (utilities) for health-related quality of life issues has become a major topic both in clinical decision making and health care planning [Pauker, 1987; Torrance, 1987]. Torrance's time tradeoff technique has been promoted as a reliable and valid method to assess numerical values, so-called utilities, for health related quality of life issues. The utility of a certain health-state is the relative value of that health-state compared with other, better and worse, health states. In the time tradeoff technique, the utility of a certain health state is assessed by asking a subject how much length of life in this health state (s)he is willing to sacrifice in order to live in a better health state for less lifetime [Torrance, 1972].

---

1 L.J.A. Stalpers, A.L.M. Verbeek and W.A.J. Van Daal. Conditionally accepted for publication in the Journal of Clinical Epidemiology
On a utility scale ranging from 0.0 for death to 1.0 for the best outcome B, the TTO test states that the utility $U(a)$ of a certain health-state A which is worse than health state B, is the proportion of the minimal amount of years $Y(b)$ a subject is willing to live in the better state compared with the (larger) amount of years $Y(a)$ in the worse state

$$U(a) = \frac{Y(b)}{Y(a)} \quad \{1\}$$

It is customary not to ask for the minimal numbers of years $Y(b)$ directly, but to assess the complement, namely maximum number of life years a subject is willing to tradeoff, from a certain length of life in the worse state A to get a better health state B.

$$Y(b) = Y(a) - \text{TRADEOFF} \quad \{2\}$$

For example, if a person is willing to tradeoff 5 from 25 years in the worse state to live a (shorter) life in the best state, the utility of the better state equals $(25-5)/25 = 0.8$.

A major assumption of the TTO is that tradeoffs are proportional, i.e. the proportion $\frac{Y(a) - \text{TRADEOFF}}{Y(a)}$ is the same for any length of life $Y(a)$ in a worse state to be traded off [Weinstein and Fineberg, 1980]. For instance, if a subject is willing to tradeoff 5 from 25 years for a better health-state, proportionality requires that the subject is willing to tradeoff the same proportion (=0.2) from any length of life in the worse state, say, 2 years from 10 years or 1 year from 5 years.

In a 1981 paper of McNeil et al., the TTO-technique was used to calculate utilities for the quality of speech in the choice between radiotherapy and surgery for laryngeal cancer [McNeil et al., 1981]. Based on this paper, we used the time tradeoff technique in medical students and general physicians to test whether proportionality of tradeoffs holds for varying lengths of life.
11.3 Population and Methods

(a) The Clinical Problem
Cancer of the vocal cords can be treated both by surgery and by radiotherapy. Survival rates following surgery are generally higher than following radiotherapy, although the differences may be small [Stalpers et al., 1987]. Successful radiotherapy generally results in a moderately hoarse but natural speech. Following surgical resection of the voice box, a patient has to learn an artificial speech. As the differences in survival rates (or the average lengths of life) between surgery and radiotherapy are only small, a definite choice between radiotherapy and surgery may depend on the value attached to quality of speech.

(b) Population
Two groups of 'medically experienced' subjects were asked to participate in an experiment to assess numerical values for the quality of speech in the choice between radiotherapy and surgery in cancer of the vocal cords. The first group consisted of 56 students of the Faculty of Public Health, age ranging from 18-36 years, mean=21 years, sex ratio male:female = 14:42. The second group consisted of 9 general physicians of the Nijmegen University Institute of General Physicians, age ranging from 39-52 years, mean=45 years, all male.

(c) The Time Tradeoff Interview
The interview began with a description of the goals and the hazards both of radiotherapy and of surgery for cancer of the vocal cords and a demonstration of a moderately hoarse 'natural speech' and of a so-called esophageal or 'artificial' speech. The subjects were asked to imagine they had cancer of the vocal cords themselves and that they had to make a choice between radiotherapy and surgery. It was suggested that the test was designed to help patients with cancer of the vocal cords to facilitate a treatment choice.

The actual test consisted of six items. In each item the subject had to assess the maximum number of life years (s)he was willing to sacrifice from a given life expectancy with an artificial speech in order to live with a natural speech. See for instance the first item:
If you receive surgery you will have 25 life years with an artificial speech. How many of these life years are you willing to sacrifice to live with a natural speech following radiotherapy?

In the six successive items the presumed length of life with an artificial speech was diminished from 25 to 20, 15, 10, 5 and 2 years. It was explicitly stated that it was also possible to tradeoff parts of a year or to express the tradeoff in months.

The test was concluded with a short questionnaire on some personal characteristics (sex, age, marital state, education, profession) and two questions investigating the ability to imagine having cancer and the ability to deal with quality of life quantitatively. These questions could be answered on a five-point scale: (a) no difficulty, (b) slight problems, (c) moderate problems, (c) severe problems and (d) practically impossible.

The TTO-test was repeated three weeks later for 15 medical students and three months later for 8 physicians.

(d) Data Analysis
Pearson's product-moment correlation coefficient $r$ was used to calculate test-retest correlations. Linear regression statistics were used to test differences in the tradeoffs among the groups (sex, age, profession, marital state) and to test the proportionality assumption. Presuming a linear model of the utility $U(a)$ of artificial speech on the length of life $L$, the condition of proportional tradeoffs holds if in the regression equation $U(a) = c + \beta L$ the coefficient $\beta = 0$ so that $c$ is a constant. Chi-square statistics were used to test differences in the answers of the two questions among the groups.

11.4 Results

On a utility scale ranging from 0.0 for death to 1.0 for a health state with a natural speech, figure 11.1 shows the utility of artificial speech as a function of the presumed length of life with artificial speech.

The broken horizontal line gives the theoretically expected graph for strict proportional tradeoffs for any length of life between 0 and 25 years. The solid line interpolates the mean utility in the 64 subjects, showing a diminishing utility attached to artificial speech for increasing life-years ($\beta = -0.003$).
Figure 11.1 - The utility of artificial speech as a function of the presumed length of life with this artificial speech. The broken line denotes the situation in which tradeoffs are proportional to length of life. The solid line denotes the mean utility values (±SD) in 56 medical students and 9 general physicians.

Table 11.1 - The ability to imagine to have cancer of the vocal cords and the ability to deal with the quality of life quantitatively in 65 subjects.

<table>
<thead>
<tr>
<th>Ability to imagine having cancer of the vocal cords</th>
<th>Ability to deal with quality of life quantitatively</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) No problems</td>
<td>3 (4.6%)</td>
</tr>
<tr>
<td>(b) Mild problems</td>
<td>5 (7.7%)</td>
</tr>
<tr>
<td>(c) Moderate</td>
<td>14 (21.5%)</td>
</tr>
<tr>
<td>(d) Severe problems</td>
<td>33 (50.8%)</td>
</tr>
<tr>
<td>(e) Practically impossible</td>
<td>10 (15.4%)</td>
</tr>
</tbody>
</table>
The graph shows a significant deviation from the proportionality hypothesis (*P* = 0.0035). No significant differences were seen between groups following stratification by sex, age, marital state or professional occupation (student vs physician), neither with respect to the average utilities nor to the extent that proportionality was violated. In 23 retested subjects, the test-retest correlation coefficient *r* for the six items ranged from 0.63 to 0.85 with *P*-values < 0.05.

Table 11.1 summarizes the answers of the two questions about the ability to imagine having cancer of the vocal cords and the ability to deal with quality of life quantitatively. A chi-square comparison did not demonstrate any significant differences for sex, age, marital state or profession, with respect to the two questions.

### 11.5 Discussion

This study confirms the findings of earlier studies that the TTO-technique produces relatively high test-retest reliability correlations (*r* = 0.63-0.85) [Torrance, 1987, Mohide et al., 1988, Churchill et al., 1987]. However, as shown in table 11.1, almost all subjects had at least moderate problems dealing with quality of life quantitatively and imagining having cancer themselves. The latter, despite the medical background of the subjects, may stress the limited value of non-patient judgments about quality of life issues for decision making both in clinical practice and in health-care politics.

This study could not identify differences in tradeoffs between the different groups, nor for sex, age, marital state or professional occupation. The latter is most remarkable as the physician group was on the average 25 years older than the student group. Both in the paper of McNeil and more recently in the paper of Mohide it is considered important to change the scale for years of future life according to the subject's age or life-expectancy [McNeil et al., 1981, Mohide et al., 1988]. This approach may enhance the feeling that the test resembles a subject's real life situation better. However, our results suggest that such an approach is not necessary since we found no significant differences in tradeoffs among the young student group and the older physician group.

The most important observation of this study is that the proportionality assumption was violated (*P* < 0.01). The subjects were willing to tradeoff a considerably bigger proportion of the length of life when faced with a long length of
life than when faced with a shorter length of life in order to live with a better quality of speech. For example, if the length of life with artificial speech was 25 years, the average utility of artificial speech would be less than 0.8, but when faced with only 5 years length of life, the same utility of artificial speech was almost 0.9. The observation that tradeoffs are not proportional can be hazardous if the TTO-utilities are used for real-life decision making. Then it is important to realize that a decision may depend on the length of life used for tradeoff in a TTO-test, and not merely on the 'real' utility. For instance, if the decision threshold for the utility of artificial speech is 0.85, i.e., everyone with a lower utility will be advised to receive radiotherapy, everyone with a higher utility will be advised to have surgery [Stalpers et al., 1986, Stalpers et al., 1988], relatively more patients will have radiotherapy if we use a TTO-test presuming a long length of life, and relatively more patients will undergo surgery if we use a test with a short length of life for tradeoff. A potential bias might have been introduced by the fact that we did not present the tradeoff test items at random but, instead, by subsequential decreasing the length of life. It is most likely that this bias leads to conservative or 'anchoring' results [Tversky & Kahneman, 1974], i.e., a subject tends to answer an item similarly to the previous item in order to seem consistent. Then, however, it is the more striking that the tradeoffs are not proportional.

The most likely explanation for violation of the proportionality presumption is that not all life years have an equal value, i.e., near years are valued relatively higher than distant life years. This explanation may be so obvious that, as far as we know, it has never been tested before. Therefore, in a tradeoff between quality and quantity of life, we have to assume an interaction of two separate utilities. A utility of the quality of life and a utility of the length of life. As long as we have no testable assumptions about how these two utilities interact, it will not be possible to separate them. A probable solution for this problem may be found by using so-called conjoint measurement techniques [Luce & Tukey, 1964]. In conjoint measurement three utility functions are made explicit. One on the quality of life, one on the length of life, and one on the combination. Presuming that the utility function of the combination of the quality and length of life results from a simple additive effect of the two separate utility functions, a so-called additive conjoint measurement test was developed to infer the three utility functions from an individually assessed preference ranking on combinations of a quality and a quantity of life [Stalpers et al., 1990]. The early experiences with additive conjoint measurement, applied to the
treatment choice in laryngeal cancer, suggest that it may be a promising alternative for Torrance's TTO-technique [Stalpers et al., 1988]. Thus establishing tradeoffs by additive conjoint measurement, we hope to meet the appeal that Mohide recently made to other researchers 'to apply, and to help develop further, this promising approach' [Mohide et al., 1988].

Acknowledgment
We kindly acknowledge the physicians from the Nijmegen University Institute of General Physicians and the Health-Science students for their voluntary participation to this study. We are obliged to Mrs. C.W. McKell for her linguistic help.

11.6 References


12. Utility assessment by additive conjoint measurement for clinical decision making

12.1 Summary

Utility theory enables the subjective evaluation of health states and life-years of medical treatments. Utility theory assumes (a) that loss in one attribute, e.g., life-years, can be compensated by gain in another attribute, e.g., quality of life, (b) that both losses and gains may be quantified, expressed as utilities in an additive model, and (c) that the treatment with the highest subjective expected utility (EU) is preferred.

In additive conjoint measurement (ACM) tradeoffs between life-years and quality of life can be derived from an individual preference ranking on pairs, each pair consisting of a combination of life-years and quality of life. ACM is based on the principle that the utility $U_{L,Q}$ for such a combination is an additive effect of separate utilities for life-years $U_L$ and quality of life $U_Q$:

$$U_{L,Q} = U_L + U_Q$$

The model holds if the preference ranking on pairs meets at least three conditions (a) transitivity, (b) independence of attributes, and (c) cancellation.

The preference ranking of combinations is assessed by having the patient make a series of choices in pairwise combinations. Thus, one can assess the extent of transitivity.

The method of measuring utilities will be illustrated by an example of a respondent who has to choose between radiotherapy and surgery for the treatment of laryngeal cancer. Radiotherapy gives a better quality of voice but also a shorter life-expectation than surgery.

12.2 Introduction

DECISION MAKING INVOLVING SURVIVAL AND QUALITY OF LIFE

Survival rates, cure rates, and life-years are among the major criteria by which treatments of patients with cancer are evaluated. In recent years, some effective methods have been developed to combine age-specific and cause-specific

---

mortality rates in predicting the prognosis in individual patients. These methods, such as the Markov model of medical prognosis, can generate an accurate description of the course of disease in individual patients and have effectively been applied to individual decision making [Beck & Pauker, 1983].

Few physicians with experience in oncology will cling to survival and cure rates as the only criteria in choosing between alternative cancer treatments. However, many physicians will hold to these rates as the main criteria of treatment success, since convenient and valid methods to measure quality of life for decision analytical purposes are still in an experimental phase.

Most of the current methods for quality of life measurement in cancer treatment, such as the Karnofsky Performance Status Scale, are predominantly descriptive and do not provide norms for guiding clinical decision making [Karnofsky et al., 1948; Mor et al., 1984; De Haes & Van Knippenberg, 1985].

Norms for the quality of life may serve for comparison with other attributes such as survival rate, cure rate or life-expectancy. The quantitative assessment of qualitative properties on a common scale is denoted utility assessment. The utility approach to medical decision making is based on the utility theory as described by Von Neumann and Morgenstern (1953). This theory states that individuals should assign (or act as if they assign) a utility to each possible outcome and faced with a set of alternatives, choose the one that yields the highest mathematical utility expectation [Hershey and Baron, 1987]. Utilities can be inferred from choices in simple cases as assessed in a simple test. These utilities can be used in the analysis of complex clinical cases. Utility assessment is based on two “tradeoff” principles:

1. Compensation: A qualitative loss on one aspect of life can be compensated by a qualitative gain on another attribute.
2. Quantification: Losses and gains can be expressed quantitatively.

The basic reference gamble and the time tradeoff technique have been used to assess utilities for clinical decision making [Pliskin et al., 1980; Torrance, 1987; McNeil et al., 1978; McNeil, 1981]. Practical and conceptual shortcomings of these methods actuate the search for alternative measurement techniques [Llwellyn-Thomas et al., 1982; Stalpers et al., 1990b].
In the present paper we describe additive conjoint measurement in the assessment of utilities both for the quality and for the quantity of life. It is demonstrated how utilities are assessed in individual patients choosing between radiotherapy and surgery for glottic cancer, and how these utilities are applied in a Markov model of medical prognosis in evaluating treatment choices.

CLINICAL PROBLEM: GLOTTIC CANCER

Patients with cancer of the vocal cords can be treated either by radiotherapy or by surgery. Radiotherapy generally consists of a course of six weeks irradiation for which a patient has to come to the treatment centre daily. Surgery consists of a resection of the involved vocal cord in patients with small tumours or complete resection of the voice box and surrounding structures in more advanced stages. In terms of cure and survival, surgery gives better results than radiotherapy, especially in more advanced stages of the disease [Stalpers et al., 1987]. In general the quality of speech is better after radiotherapy than after surgery [Dickens & Cassisi, 1983]. After successful radiotherapy, the quality of speech may range from a near normal speech to severe hoarseness, generally leaving a slightly 'cracking' voice. After surgical resection of the voice box, the quality of speech is severely impaired, and a patient has to learn some form of artificial speech. Most of these patients manage to learn a so-called oesophageal speech, whereby speech is produced by belching swallowed air. Some patients do not manage to learn an oesophageal speech, and may find relief by a so-called electronic larynx. This electronic device transforms the soundless articulation of the oral cavity to a somewhat metallic sounding speech. A minority of patients stays mute and has to communicate by writing and gesturing.

In a choice between radiotherapy and surgery for cancer of the vocal cords, it may be obvious to consider both the quality of speech and the survival or life-years as valued by the individual patient [McNeil et al., 1981].

In the first part of this paper we describe the method of utility assessment based on additive conjoint measurement (§12.3). Next we will demonstrate the use of "additive" utilities in a decision analysis using a Markov model to simulate the course of disease in patients with glottic cancer (§12.4).
12 3 Additive Conjoint Measurement (ACM)

Principles
In additive conjoint measurement (ACM) tradeoffs between life-years and quality of life are derived from an individual preference ranking on pairs of alternatives. Each pair consists of a combination of life-years (L) and quality of life (Q). It is assumed that the utility \( U_{L,Q} \) for such a combination is an additive effect of the separate utilities for life-years \( U_L \) and quality of life \( U_Q \) [Luce & Tukey, 1964]

\[ U_{L,Q} = U_L + U_Q \]  \hspace{1cm} (1)

\( U_{L,Q} \), \( U_L \) and \( U_Q \) can be inferred from a preference ranking on pairs of L and Q if at least three conditions are satisfied

1) Transitivity of preferences: If for any pairs A, B and C, you prefer pair A to B, and B to C, then you prefer A to C. The opposite (or violation) of transitivity is called circularity. For instance, if you prefer pair A to B, B to C and, C to A, there is a cycle in the preference ranking on the three elements A, B and C. In a preference ranking on the pairs A, B, C, Z, both the number of cycles and the number of pairs in each cycle contribute to the severity of intransitivity [Bezembinder, 1981]

2) Independence of attributes: The utility of one attribute is not determined by the utility of the co-attribute. For instance, let q, r, s denote the levels on attribute Q and let l, m, n denote the levels on attribute L. If you prefer the pair (q, l) to pair (r, l), you prefer q to r for any other level on L. Similarly, if you prefer (q, l) to (q, m) then you prefer l to m for any level on Q

3) Double cancellation of preferences: Let mute, esophageal and normal denote three levels on quality of speech Q, and let 3, 6 and 9 years denote three levels on life-expectancy L. Figure 12 1 gives the pairs of (Q, L) ordered in a 3x3 matrix. If you prefer the pair (n, 6) to (o, 3) (left panel) and the pair (o, 9) to (m, 6) (middle panel), then you prefer the pair (n, 9) to (m, 3) (right panel). If the condition of double cancellation is violated, there is no solution that yields additive utility functions from the preference ranking. With \( q \geq 3 \) and \( l \geq 3 \) the number of levels in Q and L respectively, the condition of double cancellation goes for any 3x3 submatrix resulting from the original qxl matrix by picking three levels on Q and three levels on L
If:

<table>
<thead>
<tr>
<th>L</th>
<th>(years)</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Q</td>
<td>3</td>
<td>6</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>mute</td>
<td>m,3</td>
<td>m,6</td>
<td>m,9</td>
<td></td>
</tr>
<tr>
<td>oesoph.</td>
<td>o,3</td>
<td>o,6</td>
<td>o,9</td>
<td></td>
</tr>
<tr>
<td>normal</td>
<td>n,3</td>
<td>n,6</td>
<td>n,9</td>
<td></td>
</tr>
</tbody>
</table>

and if:

<table>
<thead>
<tr>
<th>then:</th>
<th>(years)</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>L</td>
<td>3</td>
<td>6</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>Q</td>
<td>m,3</td>
<td>m,6</td>
<td>m,9</td>
<td></td>
</tr>
<tr>
<td>mute</td>
<td>m,6</td>
<td>m,9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>oesoph.</td>
<td>o,3</td>
<td>o,6</td>
<td>o,9</td>
<td></td>
</tr>
<tr>
<td>normal</td>
<td>n,3</td>
<td>n,6</td>
<td>n,9</td>
<td></td>
</tr>
</tbody>
</table>

Figure 12.1 • 3x3 Matrix for the pairs of Q (mute, oesophageal speech and normal speech) and L (3, 6 and 9 years).

Obtaining a preference ranking

To obtain a preference ranking on pairs, we defined five different qualities and numbers of life-years. For the quality of speech: Normal speech (N), hoarseness (H), oesophageal speech (O), electronic-larynx speech (E) and muteness (M). For life-years: 3, 6, 9, 12 and 15 years.

From five different qualities of speech and numbers of life-years, we can form 25 pairs on Q and L. Figure 12.2 gives all 25 pairs in a 5x5 matrix (n x m matrix).

<table>
<thead>
<tr>
<th>Q</th>
<th>L: 3y</th>
<th>6y</th>
<th>9y</th>
<th>12y</th>
<th>15y</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mute</td>
<td>M,3</td>
<td>M,6</td>
<td>M,9</td>
<td>M,12</td>
<td>M,15</td>
</tr>
<tr>
<td>Electronic-larynx</td>
<td>E,3</td>
<td>E,6</td>
<td>E,9</td>
<td>E,12</td>
<td>E,15</td>
</tr>
<tr>
<td>Oesophageal</td>
<td>O,3</td>
<td>O,6</td>
<td>O,9</td>
<td>O,12</td>
<td>O,15</td>
</tr>
<tr>
<td>Hoarseness</td>
<td>H,3</td>
<td>H,6</td>
<td>H,9</td>
<td>H,12</td>
<td>H,15</td>
</tr>
<tr>
<td>Normal</td>
<td>N,3</td>
<td>N,6</td>
<td>N,9</td>
<td>N,12</td>
<td>N,15</td>
</tr>
</tbody>
</table>

Figure 12.2 • 5x5 Matrix for the preference ranking on pairs of quality of speech and life-years.
For the different utilities of quality of speech $U_Q$ and life-years $U_L$, we presume a strict preference ordering:

$$U_N > U_H > U_O > U_E > U_M \quad (2a)$$

and

$$U_{15} > U_{12} > U_9 > U_6 > U_3 \quad (2b)$$

Then, 3 years mute $(M,3)$ is the least preferred pair (rank number 1) and 15 years normal speech $(N,15)$ the most preferred (rank number 25).

The preference ranking on pairs can be assessed by having the patient make a series of choices in pairwise combinations. For instance:

Which do you prefer?
A. 9 Years with a normal speech or
B. 12 Years with an oesophageal speech.

For five qualities and numbers of life-years, there are $\binom{25}{2} = 300$ different pairwise comparisons. However, since in (2a) and (2b) we assumed strict preference orders of the utilities, there are 100 comparisons which will be a dilemma to the patient and are therefore non-trivial or ‘hard’ choices. In the ACM-test we ask to assess the preference on each of the 100 dilemmas$^1$. The choices on all pairwise comparisons yield the rank number. If, for instance, one pair has been preferred 10 times over the other 24 pairs, the rank number of that pair is 10. If all rank numbers are different (from 0 to 24), the preference ranking is transitive. If the preference ranking further satisfies “independence” and “cancellation”, there generally is a perfect additive solution to the preference ranking. Several algorithms are available to find representative utility functions $U_Q$ and $U_L$ assuming additivity. We used the

\[ \frac{q^4(q-1)(l-1)}{4} \]

1. For $q$ and $l$ levels on each either attribute $Q$ and $L$, the number of unique dilemmas is:
computer programme ORDMET3 to find these additive utilities for the quality of speech and for life-years [Roskam, 1987].

**Example of ACM**

A 50-year old speech instructor, professionally involved with patients treated for laryngeal cancer, was asked to imagine having glottic cancer. To facilitate a choice between radiotherapy and surgery, we asked her to perform the ACM test. The test was performed three times with weekly intervals. Figure 12.3a gives the preference ranking ordered in a 5x5 matrix. The cell entries in the matrix represent the rank number of a pair of quality of speech (Q) and life-years (L)

<table>
<thead>
<tr>
<th>Q</th>
<th>L: 3y</th>
<th>6y</th>
<th>9y</th>
<th>12y</th>
<th>15y</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mute</td>
<td>0</td>
<td>4</td>
<td>8</td>
<td>13</td>
<td>15</td>
</tr>
<tr>
<td>Electronic larynx</td>
<td>1</td>
<td>6</td>
<td>11</td>
<td>16</td>
<td>20</td>
</tr>
<tr>
<td>Oesophageal</td>
<td>2</td>
<td>7</td>
<td>12</td>
<td>17</td>
<td>21</td>
</tr>
<tr>
<td>Hoarseness</td>
<td>3</td>
<td>9</td>
<td>14</td>
<td>19</td>
<td>23</td>
</tr>
<tr>
<td>Normal</td>
<td>5</td>
<td>10</td>
<td>18</td>
<td>22</td>
<td>24</td>
</tr>
</tbody>
</table>

Figure 12.3a • *Data matrix for the preference ranking on pairs of quality of speech and life-years as assessed by a 50-year old speech instructor ranking from 0 for the least preferred pair (3 years mute) to 24 for the most preferred (15 years normal speech).*

For this data matrix, ORDMET3 results in several possible utilities $U_Q$, $U_L$ and $U_{Q,L}$. The average is given in figure 12.3b. The solution matrix represents a tight additive transformation of the data matrix

- It is a *transformation*, as the order of the data matrix is identical to the order of the solution matrix. For example: The utility of 12 years electronic-larynx speech being 1.04, ranks as number 16 between 0.0 for 3 years mute and 1.43 for 15 years normal speech.


- It is additive, as the utility of the pair $U_L Q$ is expressed as a simple addition of the mutual independent utilities $U_Q + U_L$. For example, the utility of 12 years electronic-larynx speech is $0.18 + 0.86 = 1.04$.

- It is tight, as only marginal variations within the utilities $U_Q$ and $U_L$ preserve additivity. For example, if we decrease the utility of hoarseness with only 0.02 points, the utility of 15 years hoarseness will equally diminish from 1.30 to 1.28, and will hence become lower than the utility of 12 years normal speech being 1.29. Then, the original preference ranking is not preserved anymore in the solution matrix.

<table>
<thead>
<tr>
<th></th>
<th>L 3yr</th>
<th>6yr</th>
<th>9yr</th>
<th>12yr</th>
<th>15yr</th>
</tr>
</thead>
<tbody>
<tr>
<td>$U_L$</td>
<td>0.00</td>
<td>0.37</td>
<td>0.64</td>
<td>0.86</td>
<td>1.00</td>
</tr>
<tr>
<td>$U_Q$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Mute</th>
<th>Electronic-larynx</th>
<th>Oesophageal</th>
<th>Hoarseness</th>
<th>Normal</th>
</tr>
</thead>
<tbody>
<tr>
<td>$U_Q$</td>
<td>0.00</td>
<td>0.18</td>
<td>0.20</td>
<td>0.30</td>
<td>0.43</td>
</tr>
</tbody>
</table>

**Figure 12.3b** • Solution matrix for the additive utilities $U_Q$ and $U_L$ to the data from figure 12.3a

**Interpretation of additive utilities**

The additive utilities are close to and can, for practical purposes, be used as values on an interval scale [Roskam, 1968, p. 17-18]. The interval scales of the utilities $L$ and $Q$ have equal unities on a utility scale ranging from minus infinity to plus infinity. For practical reasons only, figure 12.3b gives a solution matrix normalized between 0.00 and 1.00. Since the scales are close to interval scales, only those transformations are allowed that preserve the ratios of intervals. For instance, the ratio of the interval between the utilities of 15 and 12 life-years, i.e., $1.00 - 0.86 = 0.14$, to the interval between the utilities of oesophageal speech and hoarseness,
\[ \text{e} \ 0.30 - 0.20 = 0.10, \text{ this ratio being } 0.14/0.10 = 1.4, \text{ has to be preserved under the transformation of the utility scales. Based on these intervals, we can say that the difference between muteness and electronic-larynx, being } 0.18, \text{ is nine times the difference between electronic-larynx speech and oesophageal speech, being } 0.02. \]

In calculating the expected utility of a treatment, additive utilities can be best interpreted in terms of gaining points per year and gaining extra points for living with a better quality of speech than being mute during that period. For instance, living 9 years will gain you 0.64 utilities. Using an OESOPHAGEAL speech during this period instead of being MUTE, will give 0.20 extra points for the quality of speech, resulting in 0.64 + 0.20 = 0.84 utilities for the combination.

We based the test on 3-year intervals, using numbers of life-years of 3 and 15 years as minimum and maximum respectively. To allow extrapolation for more or less life-years, we approximated the utilities of life-years by a logarithmic function:

\[ U_L = 0.63\ln(L) - 0.72 \quad (4) \]

As shown in figure 12.4, the data correlates well with the logarithmic function \((r = 0.99)\).

![Additive utility function of life years \(U_L\) for a speech instructor aged 50](image)

Now we can calculate the utility of any course of disease in patients with glottic cancer using the individually assessed utilities of quality of speech and quality of life.
Figure 12.5 • One of many possible courses of disease for a patient with glottic cancer

Figure 12.5 gives one of many possible courses of disease for a patient receiving radiotherapy for glottic cancer and remaining hoarse for the next two years. Then she has a tumour recurrence, receives salvage laryngectomy and remains with an oesophageal speech until she dies three years later. For the speech instructor, the utility of this specific course of disease is the sum of the utility of life-years

\[ U_L = 0.63 \ln(5) - 0.72 = 0.29 \]

and the utility of the quality of speech

\[ U_Q = \frac{2}{5} U_H + \frac{3}{5} U_O = \frac{2}{5} \times 0.30 + \frac{3}{5} \times 0.20 = 0.24 \]

resulting in the utility of the combination

\[ U_{Q,L} = U_L + U_Q = 0.29 + 0.24 = 0.53 \]

If we see the outcome of a treatment (radiotherapy or surgery) as the outcome of a stochastic process, the Von Neuman-Morgenstern theory states that the expected utility of a treatment \( EU \) is the sum of products of the probability \( P_{Q,L} \) of a certain combinations \((Q,L)\) and its utility \( U_{Q,L} \)

\[
EU = \sum_{Q=1}^{5} \sum_{L=1}^{5} (P_{Q,L} \cdot U_{Q,L})
\]  
(5)
A Markov model can be used to simulate all possible courses of disease, to calculate its utilities and to determine the contribution of each course to the expected utility of a treatment.

12.4 Decision Analysis: Treatment of Glottic Cancer

In this paragraph we will demonstrate the use of "additive" utilities in a Markov decision tree analysis for the choice between radiotherapy and surgery for glottic carcinoma, considering both the time varying mortality risks for each patient and the utilities for the quality of speech and for life-years as individually assessed by additive conjoint measurement.

DATA TREATMENT RESULTS OF GLOTTIC CANCER

<table>
<thead>
<tr>
<th>Hindi</th>
<th>Summary of a review of the literature on treatment results following radiotherapy and surgery for glottic carcino [Stalpers et al., 1987]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Radiotherapy</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Mean</strong></td>
</tr>
<tr>
<td>5-Year survival rates</td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>79%</td>
</tr>
<tr>
<td>T2</td>
<td>68%</td>
</tr>
<tr>
<td>T3</td>
<td>48%</td>
</tr>
<tr>
<td>T4</td>
<td>45%</td>
</tr>
<tr>
<td>5-Year disease-free survival rates</td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>72%</td>
</tr>
<tr>
<td>T2</td>
<td>63%</td>
</tr>
<tr>
<td>T3</td>
<td>30%</td>
</tr>
<tr>
<td>T4</td>
<td>27%</td>
</tr>
<tr>
<td>Salvage rates following tumour recurrence</td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>72%</td>
</tr>
<tr>
<td>T2</td>
<td>51%</td>
</tr>
<tr>
<td>T3</td>
<td>37%</td>
</tr>
<tr>
<td>T4</td>
<td>8%</td>
</tr>
</tbody>
</table>

N P = not published in the literature  *Mean = Mean of published reports
The Literature: In a review paper on treatment results of radiotherapy and surgery for glottic carcinoma, we described results in terms of 5-year survival, 5-year disease-free survival and salvage rate [Stalpers et al., 1987]. Table 12.1 lists the most relevant data by mode of treatment and by T-classification according to the TNM-classification of malignant tumours [UICC, 1978; AJC, 1978].

Although most investigators consider the quality of speech an important result of treatment of glottic carcinoma, only few studies report figures on treatment outcomes in terms of speech quality. Table 12.2 summarizes an overview of the literature over the past twenty years.

Table 12.2 • Quality of speech following radiotherapy and surgery of glottic cancer as reported since 1968.

<table>
<thead>
<tr>
<th>Study</th>
<th>Following radiotherapy:</th>
<th>NORMAL</th>
<th>HOARSE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>101/227 = .445</td>
<td>126/227 = .555</td>
</tr>
<tr>
<td>Stoicheff, 1975</td>
<td></td>
<td>84/110 = .764</td>
<td>26/110 = .236</td>
</tr>
<tr>
<td>Karim et al., 1983</td>
<td></td>
<td>.79-.93</td>
<td></td>
</tr>
<tr>
<td>Harwood and Rawlinson, 1983</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Study</th>
<th>Following laryngectomy:</th>
<th>OESOPHAGEAL</th>
<th>ELECTRO</th>
<th>MUTE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>37/88 = .426</td>
<td>20/88 = .227</td>
<td>31/88 = .352</td>
</tr>
<tr>
<td>King et al., 1968</td>
<td></td>
<td>.63-.84</td>
<td>.16</td>
<td>.05-.11</td>
</tr>
<tr>
<td>De Bleule and Damsté, 1972</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Retrospective Study: Data from a retrospective study in 188 patients with T₁-T₂N₀M₀ glottic carcinoma were used to model the instantaneous recurrence risk as a time-dependent probability. All patients were treated by primary radiotherapy between 1979-1987 at the Institute for Radiotherapy of the Nijmegen University Hospital, The Netherlands. Population characteristics including criteria and treat-
ment methods are described elsewhere [Van den Ende et al., 1988]. Figure 12.6 gives the actuarial disease-free survival curves stratified by T-classification. The 5-year disease-free survival rates are 75% for T1 tumours and 56% for T2 tumours.

Figure 12.7 gives the instantaneous recurrence risks per year. As shown, recurrence risks are highest in the first 2-3 years and become virtually zero after 4-6 years following radiotherapy.

Figure 12.6 • Disease-free survival following radiotherapy in T1-T2N0M0 glottic carcinoma.

Figure 12.7 • Annual tumour recurrence rate Prec per year following radiotherapy for T1-T2N0M0 glottic carcinoma.
The data of figure 12.7 were used to approximate the time-dependent model for the annual recurrence rate $P_{REC}$. Best fits were obtained by a simple linear decreasing model on time:

$$P_{REC} = a - (b)(\text{Year}) \text{ per year} \quad (6)$$

where $a$ is the initial or 'zero-time' rate and $b$ the linear regression coefficient. To avoid negative recurrence risk, we presumed that $P_{rec} = 0$ for all $P_{rec} \leq 0$. For $T_1$ and $T_2$ were found the best linear fits for:

$T_1: a = 0.096, b = 0.0134, \text{ Pearson's correlation } = -0.68$

$T_2: a = 0.39, b = 0.1, \text{ Pearson's correlation } = -0.85.$

**METHOD: STRUCTURE OF THE MARKOV MODEL**

A Markov model of prognosis with time-varying probabilities and time-varying utilities was used to analyse the decision between radiotherapy and surgery for glottic carcinoma in the context of the patient's age, tumour-stage and value judgements (utilities) of the quality of speech and the quantity of life.

We cast the problem in the framework of a Markov process to account for (1) decreasing tumour recurrence risks following radiotherapy, (2) increasing mortality rates for aging, (3) changes in the probability distribution of health states and (4) changing utilities attached to life years.

To make the model tractable, a number of assumptions was made:

1. Within each $T$-class, consecutive risks are mutually independent: A higher recurrence rate does not imply a lower or higher salvage rate.

2. The tumour recurrence rate reaches and remains zero after five years. Hence

$$P_{REC} = a - (a/5)(\text{Year}) \text{ per year} \quad (6a)$$

3. Following successful salvage treatment for a tumour recurrence, the risk of a second recurrence is assumed to be equal to the risk of a first recurrence. Figure 12.8 shows the decision tree representing the chance events in the first year following treatment for glottic carcinoma.

The Markov model is based on six transition states, in decreasing order of preference: Life with NORMAL speech, HOARSE speech, OESOPHAGEAL speech, ELECTRONIC larynx speech, MUTE or DEAD.
Figure 12.8 · *Decision tree representing the chance events in the first year following primary treatment of glottic cancer.*
Table 12.3 - Summary of the parameters used in the analysis, based on tables 1 and 2 (minimum and maximum values between brackets)

<table>
<thead>
<tr>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Recurrence rate</td>
<td>$P_{REC} = a - (a/5) \text{YEAR}$ per year*</td>
</tr>
<tr>
<td></td>
<td>$a$ Radiotherapy</td>
</tr>
<tr>
<td>$T_1$</td>
<td>0.156 (0.150-0.156)</td>
</tr>
<tr>
<td>$T_2$</td>
<td>0.216 (0.156-0.428)</td>
</tr>
<tr>
<td>$T_3$</td>
<td>0.503 (0.419-0.551)</td>
</tr>
<tr>
<td>$T_4$</td>
<td>0.538 (0.470-0.563)</td>
</tr>
<tr>
<td>$a$ Surgery</td>
<td></td>
</tr>
<tr>
<td>$T_1$</td>
<td>0.025</td>
</tr>
<tr>
<td>$T_2$</td>
<td>0.209 (0.102-0.244)</td>
</tr>
<tr>
<td>$T_3$</td>
<td>0.114 (0.209)</td>
</tr>
<tr>
<td>$T_4$</td>
<td>0.281</td>
</tr>
<tr>
<td>(2) Salvage rate</td>
<td>$P_{SALV}$</td>
</tr>
<tr>
<td>$P_{SALV}$</td>
<td></td>
</tr>
<tr>
<td>$T_1$</td>
<td>0.72</td>
</tr>
<tr>
<td>$T_2$</td>
<td>0.51</td>
</tr>
<tr>
<td>$T_3$</td>
<td>0.37 (±0.1)</td>
</tr>
<tr>
<td>$T_4$</td>
<td>0.08</td>
</tr>
<tr>
<td>(3) Probability of quality of speech</td>
<td></td>
</tr>
<tr>
<td>Following radiotherapy</td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>$P_{NORMAL}$</td>
</tr>
<tr>
<td>Hoarse</td>
<td>$P_{HOARSE}$</td>
</tr>
<tr>
<td>$P_{NORMAL}$</td>
<td>0.55</td>
</tr>
<tr>
<td>$P_{HOARSE}$</td>
<td>0.45</td>
</tr>
<tr>
<td>Following surgery</td>
<td></td>
</tr>
<tr>
<td>Oesophageal</td>
<td>$P_{OESOPH}$</td>
</tr>
<tr>
<td>$P_{OESOPH}$</td>
<td>0.70</td>
</tr>
<tr>
<td>Electronic</td>
<td>$P_{ELECTRO}$</td>
</tr>
<tr>
<td>$P_{ELECTRO}$</td>
<td>0.20</td>
</tr>
<tr>
<td>Mute</td>
<td>$P_{MUTE}$</td>
</tr>
<tr>
<td>$P_{MUTE}$</td>
<td>0.10</td>
</tr>
<tr>
<td>(4) Death from other diseases</td>
<td>$P_{ASR}$</td>
</tr>
<tr>
<td>$P_{ASR}$</td>
<td>Age specific mortality rate</td>
</tr>
<tr>
<td></td>
<td>for initial age = 50 year</td>
</tr>
</tbody>
</table>

*providing that $P_{REC} = 0$ for $\text{YEAR} \geq 5$

Following radiotherapy, a patient may remain tumour free for the rest of the year in a state with a NORMAL speech or in a state with a HOARSE speech. If there is a
tumour recurrence a patient may either be salvaged by laryngectomy, resulting in life either with OESOPHAGEAL speech or ELECTRO-larynx speech or staying MUTE for the rest of life. If a patient is not salvaged, death will follow soon (state DEAD). Patients who have been tumour-free the whole first year may undergo a similar stream of events in the next year, be it with different recurrence and mortality risks. Patients salvaged from a recurrence may remain in one of the salvaged states, or may have a second recurrence in one of the following years. Second recurrences are generally fatal within short time.

Following surgery, all patients will have some form of post-laryngectomy state (OESOPHAGEAL, ELECTRO or MUTE), both after primary treatment and after salvage treatment for a tumour recurrence.

In a Markov model, the probability distribution of states at the end of each year is the starting point for the chance events in the next year. All transitions are unidirectional. Transitions from a qualitatively inferior state to a superior state are not permitted. As it is not likely that a patient's speech quality will deteriorate other than from tumour recurrence, we did not permit transitions from NORMAL to HOARSE, from OESOPHAGEAL to ELECTRO or MUTE or from ELECTRO to MUTE. As there is no transition from the DEAD state and as all patients will eventually die, DEAD is called the Markov absorbing state. The Markov model is structured likewise in the surgical arm, with the exception that the states NORMAL and HOARSE have been replaced by OESOPHAGEAL, ELECTRO and MUTE.

ASSIGNMENT OF PROBABILITIES

The parameters used in this model are listed in Table 12.3

We modelled the instantaneous recurrence risk as a linear decreasing function of time, in accord with formula (6a). If we presume that the instantaneous recurrence rate \( P_{REC} \) becomes zero five years following primary treatment, the instantaneous recurrence rate \( P_{REC} \) and the initial recurrence rate \( a \) can be inferred reciprocally from the 5-year disease-free survival rates \( DFS_5 \) as given in Table 12.1, using formula (6a) and (7)

\[
DFS_5 = DFS_0 - \sum_{YEAR=1}^{5} P_{REC} \times DFS_{YEAR-1} \quad (7)
\]

\[
DFS_0 = 1.0
\]
Figure 12.9 gives the 5-year disease-free survival rate $DF_{S5}$ as a function of the initial recurrence rate $a$.

The probability of dying from other diseases than glottic cancer $P_{asr}$ is based on age- and sex-specific mortality rates of the Dutch population [CBS. 1987]. The baseline values of salvage rates ($P_{SALV}$) are the mean values by tumour size as described in table 12.1, with a range of ±0.1 for sensitivity analysis.

The baseline estimates for the distribution of quality of speech following radiotherapy ($P_{NORMAL}$ and $P_{HOARSE}$) and surgery ($P_{OESOPHAGEAL}$, $P_{ELECTRO}$ and $P_{MUTE}$) are mean estimates from table 12.2.

Figure 12.9 · Five-year disease-free survival rate ($DF_{S5}$) as a function of the initial rate $a$.

UTILITIES
Treatment outcomes are expressed in terms of (1) life-expectancy and (2) expected utilities adjusted for quality of speech and life-years. For illustrative purposes we used the utilities assessed in the 50-year old speech instructor given in figure 12.3b and equation (4).
12.5 Results

BASELINE ANALYSIS
Table 12.5 summarizes life-expectancies and expected utilities predicted by the Markov model under the baseline assumptions provided. Considering a life-expectancy of 31 years in women aged 50 in the Dutch population [CBS, 1987], the Markov prediction of the life-expectancy of a female patient with glottic carcinoma is reduced to 25.56 years following treatment by radiotherapy for a $T_1$ tumour to 9.09 years in a $T_4$ tumour.

Except for $T_2$ tumours, the life-expectancy following surgery is higher. In terms of expected utilities, radiotherapy is preferred in $T_1$ and $T_2$ tumours, but not in $T_3$ and $T_4$ tumours.

Table 12.5 · Baseline Analysis

<table>
<thead>
<tr>
<th></th>
<th>Radiotherapy</th>
<th>Surgery</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LE (years)</td>
<td>EU</td>
</tr>
<tr>
<td>$T_1$</td>
<td>25.56</td>
<td>1.57</td>
</tr>
<tr>
<td>$T_2$</td>
<td>22.54</td>
<td>1.41</td>
</tr>
<tr>
<td>$T_3$</td>
<td>13.54</td>
<td>0.89</td>
</tr>
<tr>
<td>$T_4$</td>
<td>9.09</td>
<td>0.63</td>
</tr>
</tbody>
</table>

LE = life-expectancy EU = expected utility

SENSITIVITY ANALYSIS
To examine how sensitive the results of the baseline analysis are for variation of the data from the literature, we performed sensitivity analysis on (1) the instantaneous recurrence risk, (2) the salvage rate, (3) and age at first treatment. As the quality of speech may be of less decisive importance in smaller tumours ($T_{1-2}$), we merely present the analysis for $T_3$ and $T_4$ tumours (see comments).
RECURRENCE RISK The instantaneous recurrence risk following surgery for T3 glottic carcinoma is solely based on one paper, presenting a 79% 5-year disease-free survival [Ennuyer and Bataini 1975] Considering that most papers on T2-tumours give worse results following surgery, the 79% in T3 tumours probably is an exceptionally favourable result If for instance we would presume an equal 5-year disease-free surgical survival in T3 tumours as the average result in T2 tumours - being 64% (see table 12 1) -, the expected utility following surgery would decrease from 1.32 to 1.14, the life-expectancy would decrease from 23.92 to 20.61 years Even under these circumstances favouring radiotherapy, surgery is to be preferred over radiotherapy for T3 glottic carcinoma Figure 12 10 gives the expected utility as a function of the 5-year disease-free survival following radiotherapy for T3 glottic carcinoma From this figure we can read that only if the 5-year disease-free survival following radiotherapy becomes higher than 35%, as described by Ennuyer and Bataini, radiotherapy is to be preferred

![Figure 12 10 · Expected utilities as a function of the 5-year disease-free survival (DFS5) following radiotherapy for T3 glottic carcinoma The horizontal reference lines indicate expected utilities following surgery presuming a favourable disease-free survival (DFS5 = 79%) and a less favourable result (DFS5 = 64%)](image)

SALVAGE RATES Salvage rates both depend on the size of the primary tumour and on the extent of the area involved in the tumour recurrence Although a recurrence following radiotherapy is generally more localized than following surgery, - as expressed by more favourable salvage rates for radiotherapy than for surgery - a recurrence from a relatively large primary tumour would both decrease the salvage rate for radiotherapy and for surgery Hence, in a sensitivity analysis on salvage rate it is not very realistic to approach radiotherapy and surgery independently
Therefore we performed a threshold analysis in which both the salvage rate for radiotherapy and surgery are varied. Figure 12.11 gives the threshold analysis on expected utilities for T3 glottic carcinoma presuming a 5-year disease-free surgical survival of 64%. From this figure we can read that for a baseline salvage rate of 0.37 for radiotherapy, radiotherapy is only preferred in the unrealistic situation that the salvage rate for surgery would be less than 0.03 instead of 0.26. Surgery will still be preferred for a variation of ± 0.1 of the baseline salvage rates. A similar threshold analysis on expected utilities for salvage rates in T4 glottic carcinoma demonstrated a preference for surgery in all realistic salvage rates for radiotherapy (not shown here).

AGE. The baseline value for age was 50 years. Most patients with glottic carcinoma are older, and as age is an important prognostic factor, we performed sensitivity analysis on age. Figure 12.12 shows the expected utilities following radiotherapy and surgery in T3 glottic carcinoma as a function of age. The analysis shows a preference for surgery for all ages, provided of course, that all other probabilities and utilities are fixed. A sensitivity analysis on age for the other T-classes did not alter the baseline conclusions (not shown here).
12.6 Conclusions

In this individual case, the analysis shows a benefit for surgery in T₃ and T₄ glottic tumours both in terms of life-expectancy and expected utility. The stability of the conclusion may depend on the representativeness both of the disease-free survival rates taken from the literature and the linear decreasing function we presumed for the instantaneous recurrence risk per year. Variation of the baseline probability of the salvage rate or age hardly influence treatment outcomes in terms of expected utility and do not change the baseline conclusions.

12.7 Comments

MODELLING ISSUES. The Markov model of prognosis, using time-dependent probabilities and time-dependent utilities, is a powerful mathematical tool to generate detailed and accurate assessments of life-expectancy and health status [Beck and Pauker, 1983]. To model the instantaneous tumour recurrence risk we used data from a retrospective study, choosing the model with the best fit, being a simple linear decreasing function of time. Elsewhere we demonstrated that 'best fit' modelling of time-dependent probabilities in a Markov process may generate accurate approximations of life-table estimates [Stalpers et al., 1990a]. A similar fitting procedure was used to approximate the utility function for life-years $u_L$. A logarithmic function showed an excellent fit with the original data.
ADDITIVE UTILITIES. Additive conjoint measurement has some practical and conceptual advantages over some other methods like the time-tradeoff test and the life standard gamble [Stalpers et al., 1988, 1990b]. The major advantage of conjoint measurement models is the possibility to assess utilities for the quality and quantity of life simultaneously in a so-called riskless context. Framing and anchoring, which are disturbing effects of other scaling methods can thus be bypassed [Tversky and Kahneman, 1974, 1981; Llewellyn-Thomas et al., 1982]. In this paper we demonstrated a practical application of ACM in medical decision making, without discussing the reliability of the utility functions. The tightness of the additive solution in figure 12.3b does not allow us to perform a simple sensitivity analysis on the utility functions $U_L$ or $U_Q$: A marginal variation of either utility function would disturb the transformation of the original preference ranking of figure 12.3a. However, figure 12.3b represents only one possible additive transformation, -the so-called centroid solution-, among a group of solutions. We may choose any of these possible solutions and examine the effect on the results in a decision tree analysis. Since we defined five utility-levels on either attribute, of which we fixed the two worst levels on 0.0 (for 3 life-years and for mute), there are eight utility levels that can vary within the limited freedom implied by the original preference ranking. In other words: We defined two attributes with eight degrees of freedom. Since it is cumbersome to perform utility analysis on a prognostic factor with eight degrees of freedom, we did not perform a sensitivity analysis on the utility functions.

CLINICAL ISSUES. In the Markov model we assumed that, -for all T-classes-, the stream of events are alike, i.e. that surgery always consists of total laryngectomy. This may not be the case in small tumours, as in $T_1$ and small $T_2$ glottic tumours, where surgery may consist of so-called conservation surgery. Conservation surgery will generally give less impairment of speech as we presumed and may hence generate higher expected utilities than we presented. However, considering the small difference in life-expectancy between surgery and radiotherapy in $T_1$ tumours and the even better predicted life-expectancy following radiotherapy in $T_2$ tumours than after surgery, radiotherapy is to be preferred in $T_1$ tumours and most $T_2$ tumours [Dickens and Cassisi, 1983; Stalpers et al, 1988a].

The utility analysis is more applicable for patients with larger glottic tumours (large $T_2$ and $T_3.4$), where both the quality and the quantity of life are of decisive
importance in the choice between radiotherapy and surgery. For these cases we demonstrated how utilities assessed by additive conjoint measurement may optimize a choice of treatment by considering a patient's personal preference concerning life-years and the quality of speech.

12.8 References


part c:
appendices
Appendix C.1: The Life Standard Gamble

Life expectancy and operation risks

Imagine that you have a serious disease for instance lung cancer, for which you can choose between surgery and radiotherapy.

If you have surgery, there is a risk of dying during or shortly after the operation. If you survive the operation, you are considered to be cured and will live for another 25 years in good health.

If you have radiotherapy, there is no risk of dying from the treatment itself. However, the disease will recur after a certain amount of years and you will eventually die from the disease. All these years you will live in good health.

By the next three questions, we want to investigate the minimal amount of certain life-years you are willing to accept following radiotherapy to avoid the risk of dying during an operation.

We only have three questions. Take your time and read them carefully: You will need the answer of the first question for the next questions. Please, do not hesitate and ask the interviewer to help you if necessary.

Situation 1:
If you have surgery you have:
• 50% chance of dying during or shortly after the operation and
• 50% chance to survive and live for 25 years in good health
If you have radiotherapy you have:
• 100% chance to live “X” years in good health

Question 1: Give the minimal amount of years “X” for which you would prefer radiotherapy to surgery:

Give these “X” years here: ....... (between 0 and 25 years)
Situation 2:
If you have surgery you have:
• 50% chance of dying during or shortly after the operation and
• 50% chance to survive and live for 25 years in good health
If you have radiotherapy you have:
• 100% chance to live “Y” years in good health

Question 2: Give the minimal amount of years “Y” for which you would prefer radiotherapy to surgery:
Give these “Y” years here: ....... (between 0 and 25 years)

Situation 3:
If you have surgery you have:
• 50% chance to survive and live “X” years in good health and
• 50% chance to survive and live for 25 years in good health
If you have radiotherapy you have:
• 100% chance to live “Z” years in good health

Question 3: Give the minimal amount of years “Z” for which you would prefer radiotherapy to surgery:
Give these “Z” years here: ....... (between “X” and 25 years)
Appendix C.2: The Time Tradeoff Test

Information on treatment of cancer of the vocal cords

Patients with cancer of the vocal cords can be treated either by surgery either by radiotherapy. Surgery involves removal of the larynx and subsequently a normal speech has become impossible. Most of these patients manage to learn a so-called artificial or "esophageal" speech. This approach involves swallowing air and then forcing it through the throat. The result is some kind of 'belching' speech. Further, surgery leaves a small opening in the neck. This breathing hole must remain open throughout the patient's lifetime but may be covered by turtleneck sweaters, scarves, etc. As a result of this hole, patients are unable to participate in water sports, which perhaps could inadvertently lead to the entrance of water in the lungs through this hole.

By radiotherapy, the voice can be preserved; therefore patients with small tumours will generally have radiotherapy. If a tumour is large, patients will have surgery, because in large tumours the survival after an operation is probably better than following radiotherapy. However, the differences are small. Therefore, it is questionable whether or not the better chances to survive outweigh the loss of normal speech.

By the following test we want to elicit the value you attach to your voice. We therefore ask you to imagine having cancer of the vocal cords and having to choose between surgery and radiotherapy.
Question 1:
If you receive surgery you will have 25 life years with an artificial speech. How many of these life years are you willing to sacrifice to live with a natural speech following radiotherapy?

I will sacrifice ...... year(s).

Question 2:
If you receive surgery you will have 20 life years with an artificial speech. How many of these life years are you willing to sacrifice to live with a natural speech following radiotherapy?

I will sacrifice ...... year(s).

Question 3:
If you receive surgery you will have 15 life years with an artificial speech. How many of these life years are you willing to sacrifice to live with a natural speech following radiotherapy?

I will sacrifice ...... year(s).

Question 4:
If you receive surgery you will have 10 life years with an artificial speech. How many of these life years are you willing to sacrifice to live with a natural speech following radiotherapy?

I will sacrifice ...... year(s).
Question 5:
If you receive surgery you will have 5 life years with an artificial speech. How many of these life years are you willing to sacrifice to live with a natural speech following radiotherapy?

I will sacrifice ....... year(s).

Personal Information

1. Date of birth:
2. Gender: male / female
3. Marital state
   O not married, no spouse
   O not married, with spouse
   O married
4. How many children do you have?
5. Highest finished education:
6. Profession:
7. How difficult is it for you to imagine having cancer of the vocal cords yourself?
   O No problems
   O Mild problems
   O Moderate
   O Severe Problems
   O Practically impossible
8. How difficult is it for you to deal with quality of life (i.e. speech) in a quantitative way?
   O No problems
   O Mild problems
   O Moderate
   O Severe Problems
   O Practically impossible
Appendix C.3: The ACM test

Which do you prefer?
A. 15 years mute

or

B. 12 years electronic larynx speech

01

Which do you prefer?
A. 15 years mute

or

B. 9 years electronic larynx speech

02

Which do you prefer?
A. 15 years mute

or

B. 6 years electronic larynx speech

03

Which do you prefer?
A. 15 years mute

or

B. 3 years electronic larynx speech

04

Which do you prefer?
A. 15 years mute

or

B. 12 years oesophageal speech

05

Which do you prefer?
A. 15 years mute

or

B. 9 years oesophageal speech

06

Which do you prefer?
A. 15 years mute

or

B. 6 years oesophageal speech

07

Which do you prefer?
A. 15 years mute

or

B. 3 years oesophageal speech

08

Which do you prefer?
A. 15 years mute

or

B. 12 years hoarseness

09

Which do you prefer?
A. 15 years mute

or

B. 9 years hoarseness

10
Which do you prefer?
A. 15 years mute
or
B. 6 years hoarseness

Which do you prefer?
A. 15 years mute
or
B. 3 years normal speech

Which do you prefer?
A. 15 years mute
or
B. 9 years electronic larynx speech

Which do you prefer?
A. 12 years mute
or
B. 6 years electronic larynx speech

Which do you prefer?
A. 12 years mute
or
B. 9 years oesophageal speech
<table>
<thead>
<tr>
<th>Question</th>
<th>A. 12 years mute</th>
<th>B. 6 years oesophageal speech</th>
</tr>
</thead>
<tbody>
<tr>
<td>Which do you prefer?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A. 12 years mute</td>
<td>or</td>
<td></td>
</tr>
<tr>
<td>or</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B. 6 years oesophageal speech</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Question</td>
<td>A. 12 years mute</td>
<td>B. 9 years normal speech</td>
</tr>
<tr>
<td>Which do you prefer?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A. 12 years mute</td>
<td>or</td>
<td></td>
</tr>
<tr>
<td>or</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B. 9 years normal speech</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Question</td>
<td>A. 12 years mute</td>
<td>B. 3 years oesophageal speech</td>
</tr>
<tr>
<td>Which do you prefer?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A. 12 years mute</td>
<td>or</td>
<td></td>
</tr>
<tr>
<td>or</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B. 3 years oesophageal speech</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Question</td>
<td>A. 12 years mute</td>
<td>B. 9 years hoarseness</td>
</tr>
<tr>
<td>Which do you prefer?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A. 12 years mute</td>
<td>or</td>
<td></td>
</tr>
<tr>
<td>or</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B. 9 years hoarseness</td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Question</td>
<td>A. 12 years mute</td>
<td>B. 6 years hoarseness</td>
</tr>
<tr>
<td>Which do you prefer?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A. 12 years mute</td>
<td>or</td>
<td></td>
</tr>
<tr>
<td>or</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B. 6 years hoarseness</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Question</td>
<td>A. 12 years mute</td>
<td>B. 6 years electronic larynx speech</td>
</tr>
<tr>
<td>Which do you prefer?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A. 12 years mute</td>
<td>or</td>
<td></td>
</tr>
<tr>
<td>or</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B. 6 years electronic larynx speech</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Question</td>
<td>A. 9 years mute</td>
<td>B. 3 years normal speech</td>
</tr>
<tr>
<td>Which do you prefer?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A. 9 years mute</td>
<td>or</td>
<td></td>
</tr>
<tr>
<td>or</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B. 3 years normal speech</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Question</td>
<td>A. 9 years mute</td>
<td>B. 3 years electronic larynx speech</td>
</tr>
<tr>
<td>Which do you prefer?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A. 9 years mute</td>
<td>or</td>
<td></td>
</tr>
<tr>
<td>or</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B. 3 years electronic larynx speech</td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Question</td>
<td>A. 9 years mute</td>
<td>B. 6 years electronic larynx speech</td>
</tr>
<tr>
<td>Which do you prefer?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A. 9 years mute</td>
<td>or</td>
<td></td>
</tr>
<tr>
<td>or</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B. 6 years electronic larynx speech</td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Question</td>
<td>A. 9 years mute</td>
<td>B. 3 years electronic larynx speech</td>
</tr>
<tr>
<td>Which do you prefer?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A. 9 years mute</td>
<td>or</td>
<td></td>
</tr>
<tr>
<td>or</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B. 3 years electronic larynx speech</td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Question</td>
<td>A. 9 years mute</td>
<td>B. 3 years electronic larynx speech</td>
</tr>
<tr>
<td>Which do you prefer?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A. 9 years mute</td>
<td>or</td>
<td></td>
</tr>
<tr>
<td>or</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B. 3 years electronic larynx speech</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>A. 9 years mute</td>
<td>or</td>
</tr>
<tr>
<td>---</td>
<td>----------------</td>
<td>----</td>
</tr>
<tr>
<td>31</td>
<td>Which do you prefer?</td>
<td>A. 9 years mute</td>
</tr>
<tr>
<td>32</td>
<td>Which do you prefer?</td>
<td>A. 9 years mute</td>
</tr>
<tr>
<td>33</td>
<td>Which do you prefer?</td>
<td>A. 9 years mute</td>
</tr>
<tr>
<td>34</td>
<td>Which do you prefer?</td>
<td>A. 9 years mute</td>
</tr>
<tr>
<td>35</td>
<td>Which do you prefer?</td>
<td>A. 9 years mute</td>
</tr>
<tr>
<td>36</td>
<td>Which do you prefer?</td>
<td>A. 6 years mute</td>
</tr>
<tr>
<td>37</td>
<td>Which do you prefer?</td>
<td>A. 6 years mute</td>
</tr>
<tr>
<td>38</td>
<td>Which do you prefer?</td>
<td>A. 6 years mute</td>
</tr>
<tr>
<td>39</td>
<td>Which do you prefer?</td>
<td>A. 6 years mute</td>
</tr>
<tr>
<td>40</td>
<td>Which do you prefer?</td>
<td>A. 6 years mute</td>
</tr>
</tbody>
</table>
Which do you prefer?
A. 15 years electronic larynx speech
   or
B. 12 years oesophageal speech

Which do you prefer?
A. 15 years electronic larynx speech
   or
B. 9 years oesophageal speech

Which do you prefer?
A. 15 years electronic larynx speech
   or
B. 6 years oesophageal speech

Which do you prefer?
A. 15 years electronic larynx speech
   or
B. 3 years oesophageal speech

Which do you prefer?
A. 15 years electronic larynx speech
   or
B. 12 years hoarseness

Which do you prefer?
A. 15 years electronic larynx speech
   or
B. 9 years hoarseness

Which do you prefer?
A. 15 years electronic larynx speech
   or
B. 6 years hoarseness

Which do you prefer?
A. 15 years electronic larynx speech
   or
B. 3 years hoarseness

Which do you prefer?
A. 15 years electronic larynx speech
   or
B. 12 years normal speech

Which do you prefer?
A. 15 years electronic larynx speech
   or
B. 9 years normal speech
Which do you prefer?
A. 15 years electronic larynx speech

or

B. 6 years normal speech

Which do you prefer?
A. 15 years electronic larynx speech

or

B. 3 years normal speech

Which do you prefer?
A. 12 years electronic larynx speech

or

B. 9 years oesophageal speech

Which do you prefer?
A. 12 years electronic larynx speech

or

B. 6 years oesophageal speech

Which do you prefer?
A. 12 years electronic larynx speech

or

B. 3 years oesophageal speech

Which do you prefer?
A. 12 years electronic larynx speech

or

B. 9 years hoarseness

Which do you prefer?
A. 12 years electronic larynx speech

or

B. 6 years hoarseness

Which do you prefer?
A. 12 years electronic larynx speech

or

B. 3 years hoarseness

Which do you prefer?
A. 12 years electronic larynx speech

or

B. 9 years normal speech

Which do you prefer?
A. 12 years electronic larynx speech

or

B. 6 years normal speech
Which do you prefer?
A. 12 years electronic larynx speech
   or
B. 3 years normal speech

Which do you prefer?
A. 9 years electronic larynx speech
   or
B. 6 years oesophageal speech

Which do you prefer?
A. 9 years electronic larynx speech
   or
B. 6 years oesophageal speech

Which do you prefer?
A. 9 years electronic larynx speech
   or
B. 3 years oesophageal speech

Which do you prefer?
A. 9 years electronic larynx speech
   or
B. 3 years hoarseness

Which do you prefer?
A. 9 years electronic larynx speech
   or
B. 3 years normal speech

Which do you prefer?
A. 9 years electronic larynx speech
   or
B. 3 years normal speech

Which do you prefer?
A. 6 years electronic larynx speech
   or
B. 3 years hoarseness

Which do you prefer?
A. 6 years electronic larynx speech
   or
B. 3 years normal speech
Which do you prefer?
A. 15 years oesophageal speech
   or
B. 12 years hoarseness

Which do you prefer?
A. 15 years oesophageal speech
   or
B. 9 years normal speech

Which do you prefer?
A. 15 years oesophageal speech
   or
B. 6 years normal speech

Which do you prefer?
A. 15 years oesophageal speech
   or
B. 3 years normal speech

Which do you prefer?
A. 12 years oesophageal speech
   or
B. 9 years hoarseness

Which do you prefer?
A. 12 years oesophageal speech
   or
B. 6 years hoarseness
Which do you prefer?
A. 12 years oesophageal speech
   or
B. 3 years hoarseness

Which do you prefer?
A. 12 years oesophageal speech
   or
B. 9 years normal speech

Which do you prefer?
A. 12 years oesophageal speech
   or
B. 6 years normal speech

Which do you prefer?
A. 9 years oesophageal speech
   or
B. 6 years normal speech

Which do you prefer?
A. 9 years oesophageal speech
   or
B. 3 years normal speech

Which do you prefer?
A. 6 years oesophageal speech
   or
B. 3 years normal speech
Which do you prefer?
A. 15 years hoarseness
or
B. 12 years normal speech

Which do you prefer?
A. 15 years hoarseness
or
B. 12 years normal speech

Which do you prefer?
A. 15 years hoarseness
or
B. 9 years normal speech

Which do you prefer?
A. 15 years hoarseness
or
B. 6 years normal speech

Which do you prefer?
A. 15 years hoarseness
or
B. 3 years normal speech

Which do you prefer?
A. 12 years hoarseness
or
B. 6 years normal speech

Which do you prefer?
A. 12 years hoarseness
or
B. 3 years normal speech

Which do you prefer?
A. 12 years hoarseness
or
B. 9 years normal speech

Which do you prefer?
A. 12 years hoarseness
or
B. 6 years normal speech

Which do you prefer?
A. 12 years hoarseness
or
B. 3 years normal speech

Which do you prefer?
A. 6 years hoarseness
or
B. 3 years normal speech
Appendix C.4: The FORTRAN ACM-Programme

PROGRAM ACM
C A FORTRAN programme to assess the preference ranking of pairs of life expectancy and quality of life from pairwise comparisons. The programme generates the results of three observations, producing a fourth 'averaged' ranking for which the violation of transitivity of the preference ranking is assessed.
C All variables are integers. There are two attributes: Year (Q) and quality of speech (R) with M and N levels respectively. Y is the first pair of a pairwise comparison composed by a year-level (YQ) and a quality level (YR). X is the second pair composed by XQ and XR.

CHARACTER*5 IR, ANTW
INTEGER*2 M, N, Y, X, YQ, YR, XQ, XR, Z
C The number of levels on each attribute is limited to five.
DIMENSION MTX(25,25,6), NSOM(25), MSOM(25), ISOM(25)
DIMENSION INDEX(25), MY(25), MX(25), NSOMRIJ(25), NRIJ(25)
DIMENSION IQ(5), IX(25), NCUM(25), NPAIRS(25)
C First read the number of levels on each attribute.
PRINT '( How many life expectancies are there (max 5)? ', $)
READ(5, *) M
PRINT '( How many life qualities are there (max 5)? ', $)
READ(5, *) N
PRINT '( There are ', 12, ' life expectancies and ', 12, ' life qualities'), M, N
WRITE(6, $)
DO 5 YQ = 1, M
PRINT '( Give the value of life expectancy ', 12, ' ', $)', YQ
READ(5, *) IQ(YQ)
5 CONTINUE
WRITE(6, $)
DO 10 YR = 1, N
PRINT '( Give a name (1 capital) to life quality ', 12, ' ', $)', YR
READ(5, 2005) IR(YR)
10 CONTINUE
Z = 0
IOUT = 6
Z = Z + 1
IF (Z .EQ. 4) THEN
C Calculate the cumulative result of three observations
   DO 14 Y=1,M*N
   DO 14 X=1,M*N
   ΜΤΧ(Υ,Χ,Ζ) = MTX(Y,X,1) + MTX(Y,X,2) + MTX(Y,X,3)
14   CONTINUE
   GO TO 401
ELSE IF (Ζ .EQ. 5) THEN
C Calculate the cumulative result of three observations,
C and transform them into a binary notation
   DO 15 Y=1,M*N
   DO 15 X=1,M*N
   ΜΤΧ(Υ,Χ,Ζ) = ΜΤΧ(Υ,Χ,Ζ-Ι)
   IF (ΜΤΧ(Υ,Χ,Ζ) .GT. DTHEN
      ΜΤΧ(Υ,Χ,Ζ) = 1
   ELSE
      ΜΤΧ(Υ,Χ,Ζ) = 0
   END IF
15   CONTINUE
   GO TO 401
ELSE IF (Ζ .EQ. 6) THEN
   STOP
END IF
WRITE(5,*) 'Results of replication no. ',Z
C Is the paired comparison a dilemma or not? KTEL is the counter of dilemmas
   KTEL=0
   DO 200 YQ=1,M
   DO 200 YR=1,N
   DO 200 XQ=1,M
   DO 200 XR=1,N
   Y = YR + N*(YQ-1)
   X = XR + N*(XQ-1)
   IF (YR .LT. XR) GO TO 20
      GO TO 25
20   IF (YQ .GT. XQ) GO TO 125
MTX(Y,X,Z) = 1
GO TO 200

25   IF (YR .GT. XR) GO TO 35
     GO TO 45

35   IF (YQ .LT. XQ) GO TO 200
     MTX(Y,X,Z) = 1
     GO TO 200

45   IF (YQ .LE. XQ) MTX(Y,X,Z) = 0
     IF (YQ .GT. XQ) MTX(Y,X,Z) = 1
     GO TO 200

C Read the binary value of the dilemma:

125  KTEL = KTEL + 1
     WRITE(6,3001) KTEL, IQ(YQ), IR(YR), IQ(XQ), IR(XR)

3001  FORMAT(I3,' Give the value (1,0) to',I2,A2,' ;', I2,A1,' $')
     READ(5,*) MTX(Y,X,Z)

C Give the complementary binary value to the equivalent (trivial) dilemma

200  CONTINUE
     MN = M*N

C Titles to the results

401  IF (Z .LT. 4) THEN
     WRITE(IOUT,2012) Z
     ELSE IF (Z .EQ. 4) THEN
     WRITE(IOUT,2009) 'Cumulative results'
     ELSE IF (Z .EQ. 5) THEN
     WRITE(IOUT,2009) 'Cumulative results (binary)'
     END IF
     WRITE(IOUT,*) 'Unsorted matrix of',M,' life expectancies and',N,' life qualities'
     WRITE(IOUT,2010) (IQ((X-1+N)/N), IR(MOD(X-1,N)+1), X= 1,MN)
     WRITE(IOUT,2007) (' ', X = 1,MN+2)

C Sum rows

DO 260 Y = 1,MN
     NSOM(Y) = 0
DO 255 X = 1,MN
     NSOM(Y) = NSOM(Y) + MTX(Y,X,Z)
C Write out the matrix by row
  DO 280 YQ = 1,M
  DO 275 YR = 1,N
    DO 270 XQ = 1,M
    DO 265 XR = 1,N
    Y = YR + N*(YQ-1)
    X = XR + N*(XQ-1)
  265 CONTINUE
  270 CONTINUE
WRITE(IOUT,2001) IQ(YQ), IR(YR), (MTX(Y,X,Z), X = 1,MN), NSOM(Y)
  275 CONTINUE
  280 CONTINUE
WRITE(IOUT,2007) (' ', X = 1,MN+2)
C Sum columns and write the sum-score below the matrix
  DO 325 X = 1,MN
  MSOM(X) = 0
  DO 320 Y = 1,MN
    MSOM(X) = MSOM(X) + MTX(Y,X,Z)
  320 CONTINUE
  325 CONTINUE
WRITE(IOUT,2002) '-', (MSOM(X), X = 1,MN), '--'
  WRITE(IOUT,*),
C Prepare matrix for sorting by marginal row and column sum scores
  WRITE(IOUT,2009) 'Matrix sorted by marginal sum scores:'
C Step 1: Assignment of an index to each row and column
  DO 330 K = 1,MN
  ISOM(K) = NSOM(K)
  330 CONTINUE
  DO 350 K = 1,MN
    KL = ISOM(K)
    NKL = K
    DO 340 L = 1,MN
      IF(KL .LE. ISOM(L)) GOTO 340
  340 CONTINUE
NKL = L
KL = ISOM(L)]

CONTINUE
INDEX(K) = NKL
ISOM(NKL) = 9999

CONTINUE
C Step 2: Write the names of each pair sorted by column sum-score
WRITE(IOUT,2010) (IQ((INDEX(J)-1+N)/N),
1 IR(MOD(INDEX(J)-1,N)+1), J=1,MN)
WRITE(IOUT,2007) (' ', X=1,MN+2)
C Step 3: Sort the rows and columns marginal sum scores using a
C row-index (INDEX(I)) and a column-index (INDEX(J))
DO 370 I=1,MN
K=INDEX(I)
KQ=(K-1+N)/N
KR=MOD(K-1,N)+1
WRITE(IOUT,2001) IQ(KQ), IR(KR), (MTX(K,INDEX(J),Z), J=1,MN), NSOM(K)
CONTINUE
WRITE(IOUT,2007) ('_____', X=1,MN+2)
WRITE(IOUT,2002) '-', (MSOM(INDEX(l)), l=1,MN), '--'
C Write the preference ranking
WRITE(IOUT,*)' ' ',
WRITE(IOUT,2009) 'Preference ranking of pairs:'
WRITE(IOUT,2006) (IQ((INDEX(I)-1+N), IR(MOD(INDEX(I)-1,N)+1), I=1,MN
C Write the corresponding row sum score
WRITE(IOUT,2008) (NSOM(INDEX(I)), I = 1,MN)
C Calculate and write the cumulative row sum score
NCUM(INDEX(1)) = NSOM(INDEX(1))
DO 510 I=2,MN
NCUM(INDEX(I)) = NCUM(INDEX(I-1)) + NSOM(INDEX(I))
CONTINUE
WRITE(IOUT,2008) (NCUM(INDEX(I)), I = 1,MN)
C Calculate and write the corresponding pair-value
DO 520 I = 1,MN
NPAAR(I) = (I*(I-1))/2
CONTINUE
WRITE(IOUT,2008) (NPAAR(I), I = 1,MN)
C Write a N*M preference matrix of pairs of life expectancy and life quality
WRITE(IOUT,2029) (IR(YR), YR = 1,N)
WRITE(IOUT,2028) (' ', X=1,N+2)
DO 530 YQ=1,M
WRITE(IOUT,2030) IQ(YQ), (NSOM(I), I=((YQ*N)-N+1), YQ*N)
CONTINUE
WRITE(IOUT,2028) (' ', X=1,N+2)
C Do you want the results to be filed ?
IF (IOUT.EQ. 9) GO TO 12
WRITE(6,*)
PRINT '(" Do you want these results to be filed ? ")'
READ(5,2005) ANTW
IF (ANTW.EQ. 'J') THEN
  IOUT = 9
  OPEN (UNIT=9, FILE='ACM.DAT', STATUS='UNKNOWN',
     1 CARRIAGECONTROL='FORTRAN')
  GO TO 401
ELSE
  GO TO 13
ENDIF
END
Samenvatting

Dit proefschrift geeft de resultaten van een onderzoek naar de toepassing van klinische besliskunde ter rationalisering van behandlingskeuzen voor en van patiënten met kanker.

Deel A geeft een inleiding tot de klinische besliskunde in de oncologie; deel B geeft een meer gedetailleerde beschrijving van klinische besliskunde bij patiënten met een larynxcarcinoom of een mondholtecarcinoom.

Zoals wordt beschreven in hoofdstuk A.1 wordt een formele beslissingsanalyse bij een patiënt met kanker vooral bemoeilijkt door (a) een gebrekkige kennis van relevante kansen en (b) de moeilijkheden bij het kwantificeren van de kwaliteit van het leven op waarderingsschalen uitgedrukt in utiliteiten. Een bijzonder probleem binnen de oncologische praktijk is (c) de tijdsafhankelijkheid van de kansen en de utiliteiten.

In hoofdstuk A.2 wordt de beslisboom beschreven voor het structureren en evalueren van klinische keuzeproblemen, rekening houdend met de onzekerheid van de geschatte kansen en met meerdere met elkaar wedijverende uitkomsten, zoals de levensduur en de kwaliteit van het leven. In een gevoeligheids- of sensitiviteitsanalyse wordt de stabiliteit van een keuze nagegaan door het variëren van de uitgangswaarden in de beslisboom binnen de waargenomen spreiding van kansen. Middels een sensitiviteitsanalyse kan worden aangegeven wanneer er sprake is van een beslissingsprobleem en wanneer het van doorslaggevende betekenis is om andere factoren, -in het bijzonder de kwaliteit van het leven, uitgedrukt in utiliteiten- , nauwkeurig te bepalen. Op basis van een literatuurstudie, beschreven in hoofdstuk B.6, wordt in hoofdstuk B.7 een beslisboomanalyse beschreven voor de keuze tussen radiotherapie en chirurgie bij patiënten met een T2N0M0 glottisch carcinoom.

In hoofdstuk A.3 wordt, gebaseerd op de door Beck e.a. in 1982 beschreven DEALE-methode, een eenvoudige methode beschreven om de levensverwachting, het gemiddelde sterfsterfisco per jaar en de vijf-jaars overleving te schatten op basis van een combinatie van leeftijdsspecifieke en ziektespecifieke sterftecijfers.

Echter, noch de eenvoudige beslisboom, noch de DEALE lenen zich goed voor het structureren van het voor kanker typerende tijdsafhankelijke ziektebeloop. Het in hoofdstuk A.4 beschreven Markov model is een simpele methode om tijds-
Samenvatting

afhankelijke risico's en met de tijd variërende utiliteiten te modelleren en te evalueren. Gebruik makend van actuarieel beschreven incidenties en sterftecijfers, zoals beschreven in hoofdstukken B 8 en B 9, wordt in hoofdstuk B 10 het Markov model gebruikt voor het schatten van de effectiviteit en de haalbaarheid van een gerandomiseerde studie naar de waarde van bronchoscopie voor de vroege opsporing van longkanker bij patienten met een larynxcarcinoom.

Subjectieve waardeoordelen van medische behandelingen, uitgedrukt als utiliteiten zijn tot nog toe slechts beperkt toepasbaar gebleken in de oncologische praktijk. In hoofdstuk A 5 worden de resultaten beschreven van onderzoek met een drietal methoden om utiliteiten te meten.

De in hoofdstuk A 5.2 beschreven 'life standard gamble' is alleen al wegens de grote praktische moeilijkheden die proefpersonen ondervinden ongeschikt voor klinische toepassingen.

De in hoofdstuk A 5.3 beschreven 'time tradeoff technique' is niet valide omdat een belangrijke voorwaarde voor deze methode, namelijk de proportionaliteit van tradeoffs voor alle gegeven levensduren, in hoge mate geschonden wordt door de meeste proefpersonen. Hoofdstuk B 11 geeft de resultaten van de time tradeoff test bij een 65-tal medisch geschoolde proefpersonen.

Het in hoofdstuk A 5.4 beschreven additief conjunct meetmodel (ACM) voor het meten van utiliteiten komt tegemoet aan een aantal praktische en conceptuele bezwaren van de eerder beschreven methoden. Gegeven een aantal modelvoorwaarden kunnen utiliteitsfuncties voor de kwaliteit van het leven en voor de levensduur worden afgeleid uit een individueel bepaalde voorkeursvolgorde van paren van kwaliteit en duur. In hoofdstuk B 12 wordt beschreven hoe de middels de ACM-methode bepaalde utiliteiten in combinatie met het Markov model gebruikt kunnen worden ter bepaling van de keuze tussen radiotherapie en chirurgie van het larynxcarcinoom, rekening houdend met de individuele omstandigheden en voorkeuren van de patiënt.
Stellingen

behorende bij het proefschrift
‘clinical decision making in oncology’

1. De grondslagen van de klinische besliskunde en de daarop gebaseerde technieken bieden voldoende basis voor een verantwoorde besluitvorming bij de keuze van een behandeling door en voor een patiënt met kanker, rekening houdend met de individuele omstandigheden en persoonlijke voorkeuren van de patiënt.

2. De “life standard gamble” en “Torrance’s time tradeoff techniek” zijn ongeschikt voor het bepalen van individuele afwegingen tussen de kwaliteit van het leven en de levensduur.

3. Het jaarlijks röntgenologisch longonderzoek in de nacontrole van patiënten met een mondholtecarcinoom is niet zinvol na een ziektevrije periode van twee jaar.

4. Het levensreddende effect van regelmatige bronchoscopie in de nacontrole van patiënten met een larynxcarcinoom is niet aantoonbaar.

5. De levensverwachting is een slechte maatstaf voor het succes van een behandeling als een patiënt toch niets meer van het leven verwacht.

6. De slechtere prognose van patiënten met een supraglottisch larynxcarcinoom vergeleken met die van patiënten met een glottisch carcinoom kan vrijwel volledig worden toegeschreven aan de hogere sterfte aan een longmaligniteit na het supraglottisch carcinoom.

8. Een wetenschappelijke theorie is slechts een model dat wij opstellen om onze waarnemingen te beschrijven: het bestaat alleen in onze geest. Dus is het zinloos ons af te vragen: wat is echt? Het komt er op aan welk model de meest bruikbare beschrijving oplevert. 
(Stephen Hawking, Het Heelal 1988)

(B.E. de Pauw, 22 januari 1987)

10. Men mag stellen dat de vraag ‘ben ik mijns broeders hoeder?’ bevestigend beantwoord kan worden, maar dat daarbij wel aangetekend dient te worden dat mijn broeder beschouwd moet worden als een volwassen mens. 

11. Prospectief onderzoek leidt tot bewijs, retrospectief onderzoek tot nadenken.

12. Epidemiology: The art of destroying your conclusions.

Nijmegen, 9 april 1991

Lukas Stalpers