ASSESSMENT OF PROGNOSTIC FACTORS IN CLINICAL MEDICINE

W.C.J. HOP

Printed by: Haveka B.V., Alblasserdam, The Netherlands.

Aan Annette,
Walter, Anne-Wil

ASSESSMENT OF PROGNOSTIC FACTORS IN CLINICAL MEDICINE

Evaluatie van Prognostische Factoren in de Klinische Geneeskunde

PROEFSCHRIFT

Ter verkrijging van de graad van doctor aan
de Erasmus Universiteit Rotterdam op gezag van de
Rector Magnificus
Prof. Dr P.W.C. Akkermans M.A.
en volgens besluit van het College voor Promoties

De openbare verdediging zal plaatsvinden op
woensdag 25 september 1996 om 13:45 uur

door

Willebrordus Cornelis Johannes Hop

geboren te Made

Promotiecommissie

Promotor:	Prof. R. van Strik
Overige leden:	Prof. Dr J.D.F. Habbema Prof. Dr S.W. Schalm Dr Th. Stijnen

CONTENTS page

CHAPTER 1

INTRODUCTION

Patients suffering from a certain disease generally differ greatly regarding the course of their disease. The identification of factors from which much of this variability between patients can be explained may be of great importance. The study of such prognostic factors in clinical medicine is usually done for a variety of reasons which to some extent may overlap each other. One of the reasons may be to better understand how the disease is likely to behave. It is hardly ever possible nowadays to study the pure natural history of a disease because usually some form of treatment will have been undertaken. The treatment often will be curative in intent, but may also be directed towards the relief of symptoms or avoiding or delaying these in case no curative treatment is available. It is known from their natural history that certain illnesses do not require treatment at all because they are self-limited processes. Knowledge about prognostic factors in such cases will make it possible to provide information to the patient about the probable duration and course of the disease. Knowledge of factors which are related to the outcome of disease may be helpful to physicians in understanding the mechanism of the disease. Such better understanding may assist in modifying treatment protocols, searching for new treatments or formulating strategies for the optimal use of expensive medical tests. Sometimes knowledge of prognostic factors may be of use to alter the course of disease by remedial action. This will particularly be so if these factors are elements of lifestyle such as diet, exercise or habits. For instance, patients with alcoholic liver cirrhosis who stop taking alcohol appear to have a better prognosis regarding survival as compared to those who continue drinking alcohol [Borowsky et al, 1981]. Such scientifically based facts may be helpful in encouraging patients to change their drinking habits.

Knowledge of prognostic factors often will be of importance in planning therapeutic trials. It may be desirable to evaluate certain therapies only in patients who belong to certain prognostic categories. Also the number of patients required in a randomized clinical trial depends on the distribution of prognostic factors among patients to be included in the study.

A fundamental dictum of all scientific experimentation is that in analysing experiments one has to account for all known factors which may influence the results. To adhere to this principle in randomized clinical trials it is usual to stratify on a few important prognostic factors, particularly if the size of trial is relatively small. This will guarantee that the treatment groups will be comparable regarding these factors. In the phase of analysis additional other factors may be taken into account to obtain more precise estimates of the treatment effects. Non-randomized studies to compare interventions, e.g. studies with historical control groups, can be of value only when all major prognostic factors are known and taken into account at analysis [Armitage & Gehan, 1974]. In analysing such studies it is essential that comparisons are made within groups of patients who are comparable with respect to such factors. Nevertheless, great care is still needed in the interpretation of the outcomes of such studies [Byar, 1980]. In this respect it is worth mentioning that the use of statistical tests to assess the comparability of groups is improper [Dales and Ury, 1978]. Large imbalances in unimportant variables will not matter much in studies using historical controls, but

even small imbalances in important ones may severely bias treatment comparisons. Although in randomized studies the distribution of prognostic factors will tend to be equal in the various randomized groups, it is nevertheless of great value to take account of them in the analysis. The treatment effects will be estimated with much greater precision if the various known sources of variation are adequately taken account of in the analysis. In the last decades, sophisticated and powerful statistical models have been developed which deal with this issue.

A major challenge in analyzing clinical studies is to recognize the possibility that some patient subgroups may benefit more from a certain treatment than others. However, searching for such so-called treatment-factor interactions raises great statistical problems though. Because of the usually large number of ways the patients studied can be divided into subgroups, it is likely that by chance alone some apparent treatment-factor interactions will arise. Special caution is required in drawing conclusions from such findings [Peto et al, 1976]. The general solution to this problem is to limit the special comparisons to only a few prespecified subgroups.

There are many kinds of prognostic factors depending on the type of the disease studied. In the field of cancer Byar (1983) classified these into three categories: hostfactors (e.g. age, comorbidity), tumour characteristics (e.g. histopathological parameters, extent of surgical removal of the tumour), and effects of the tumour on the host (e.g. weightloss, performance status).

For a variety of diseases there are formal disease classification systems which are strongly related to prognosis. Well known systems for instance are the TNM-classification for cancer patients [UICC, 1992], the Child-Pugh classification for liver disease [Conn, 1981] and the APACHE-score [Knaus et al, 1986] for use in intensive care medicine.

The TNM-classification system is a system based upon experts opinions and is based on three separate entities: the anatomical extent of the primary tumor (T), the condition of the regional lympnodes (N) and the presence/absence of metastases(M). Combinations of these three components lead to a prognostic stage grouping of the cancer concerned.

The Child-Pugh score is based upon clinical experience and involves three biochemical variables (serum bilirubin and albumin level, and prothrombin time) and the two clinical characteristics absence/presence of ascites and encephalopathy. Each of these variabeles is given one to three points, leading to a total score with a minimum of 5 and a maximum of 15 points, with the latter indicating a very bad liver function and poor prognosis. Although originally utilized to predict prognosis after surgical treatment [Pugh et al, 1973], the classification has proven to be useful for patients with liver cirrhosis in general.

The Apache-score combines many variables, mostly physiological, by attaching a weight to each of the separate items. The score, when evaluated at entry into the intensive care unit, is capable of providing probabilities of dying on the intensive care unit. The weights of the various features in this prognostic index have been derived using multiple logistic regression [Cox, 1970].

The description of the patterns of events, e.g. relapse or death, along time under specified conditions, and the relation thereof with various factors such as manifestations of disease, characteristics of the patients or interventions, is a major goal in many clinical studies. The question of prognosis need not solely be asked at the time of diagnosis. Also an update of prognosis during follow-up on the basis of the latest

observations on patients will be valuable. Prognostic estimates are expected to be most accurate when based on a careful analysis of data of sufficient good quality. When clinical data have been gathered rigorously, preferably in a prospective manner, adequate statistical methods are required to derive optimal estimates of patient outcomes. The statistical aspects of the identification and evaluation of prognostic factors is the subject of the present thesis.
In the following chapters frequent use is made of the terms death and failure to denote the event determining prognosis. The methods described, however, will equally be useful in case the event considered is favourable, e.g. the achievement of remission of disease.

References

Armitage P, EA Gehan. Statistical methods for the identification and use of prognostic factors. Int J Cancer 1974; 13: 16-36.

Borowski SA, S Strome, E. Lott. Continued heavy drinking and survival in alcoholic cirrhosis. Gastroenterology 1981; 80: 1405-1409.

Byar DP. Why databases should not replace randomized clinical trials. Biometrics 1980; 36: 337-342.

Byar DP. Identification of prognostic factors. In: Cancer Clinical Trials. (Eds. Buyse ME, MJ Staquet, RJ Sylvester). 1984, Oxford University Press , Oxford.

Cox DR. The analysis of binary data. 1970. Methuan, London.

Conn HO. A peek at the Child-Turcotte classification. Hepatology 1981; 6: 673-676.

Dales LG, HK Ury. An improper use of statistical significance testing in studying covariables. Int J Epidemiol 1978; 7: 373-375.

Knaus WA, EA Draper, DP Wagner, JE Zimmerman. APACHE II: a severity of disease classification system. Crit Care Med 1986; 13: 818-829.

Peto R, MC Pike, P Armitage, NE Breslow, DR Cox, SV Howard, N Mantel, K McPherson, J Peto, PG Smith. Design and analysis of randomised clinical trials requiring prolonged observation on each patient. Br J Cancer 1976; 34: 585-612.

Pugh RNH, IM Murray-Lyon, JL Dawson, et al. Transection of the esophagus for bleeding esophageal varices. Br J Surg 1973; 60: 646-649.

TNM atlas. International Union Against Cancer. (Eds Spiessl B, et all). 1992, Springer Verlag, Heidelberg.

CHAPTER 2

EVENT HISTORY ANALYSIS

The prognosis of an individual patient can be characterized by whether or not a properly defined event of interest will occur, e.g. death, and the time it takes to do so. Unfortunately, any prognostic statement about individual patients will nearly always be imprecise and, usually large, uncertainties will remain. Practice demonstrates that predictions of future events can only be made in probabilistic terms in the vast majority of cases. In this chapter statistical methods are reviewed to evaluate patient characteristics regarding their ability to reduce probabilistic uncertainties in estimating prognosis.

1. Characterisation of event occurrence.

If for an individual patient from some population, T denotes the time until the occurrence of the event, e.g. death, measured from some time "zero", there are several ways to describe the pattern of events along time for the population. Time "zero" is defined as the time corresponding to the entrance into the population, e.g. diagnosis, initiation of treatment, the timepoint at which remission of disease was established, etcetera. Time "zero" (t=0) marks the point in time from which prognosis is to be made. A descriptor of the pattern of events along time in a population which is most often used is the "survival function" S(t). This function denotes at each time point t the fraction of the patients in the population which did not experience the event at or before that time. Expressed in another way, it gives for an arbitrary subject in the population the probability that T exceeds t, i.e.

$$S(t) = Pr\{T > t\} \tag{1.1}$$

Another measure, which in practice is seldom used and mainly serves in theoretical developments, is the probability density function f(t), which is defined as:

$$f(t) = -\frac{d\ S(t)}{dt} \tag{1.2}$$

The quantity f(t)dt indicates the unconditional probability for the event to occur within the small interval (t,t+dt). A third measure to characterize the pattern of events along time is the hazard rate function $\lambda(t)$. This measure, depending on the definition of the event, is also called failure rate, death rate or incidence rate. The hazard rate function is defined as:

$$\lambda(t) = -\frac{d\ S(t)/dt}{S(t)} \tag{1.3}$$

$\lambda(t)$ can be interpreted as a conditional probability density function of failure: $\lambda(t)dt$ gives the probability of the event to occur in the small interval (t,t+dt) among those patients still at risk for the event at time t. The hazard rate function in many cases is

more informative than either the probability density function f(t) or the survival function S(t).
There are a number of relations between $\lambda(t)$, f(t) and S(t), because either one of them determines the other two:

$$\lambda(t)=\frac{f(t)}{S(t)} \tag{1.4}$$

and

$$\lambda(t)=-d \log S(t)/dt \tag{1.5}$$

in which log denotes the natural logarithm,
and

$$S(t)=e^{-\Lambda(t)} \tag{1.6}$$

in which $\Lambda(t)$ denotes the cumulative hazard $\int_0^t\lambda(\tau)d\tau$.
As $\lambda(t)$ represents the rate of failure among that part of the population still at risk for the event at time t, it is better capable of describing the intensity of the failure process at time t as compared to the survival function S(t). In some cases it is possible on biological grounds to predict whether the hazard rate will be increasing or decreasing along time. For instance, in the period shortly after a surgical hernia repair one might expect that the instantaneous recurrence rate will decrease due to healing and strengthening of tissues. After some time, when the patients resume their daily activities, possibly including hard labour, a rise of the relapse rate may be anticipated.
The survivor function S(t) is related to $\lambda(t)$ only through its cumulative function $\Lambda(t)$ (see 1.6). As is evident from relation (1.5), however, much of the information about $\lambda(t)$ can be read from graphing the survival function S(t) on a logarithmic scale. If the slope of the curve thus obtained increases along some time interval of the time axis, the instantaneous failure rate $\lambda(t)$ will also increase during the same interval. An interval on which the slope of the logarithmically transformed survival function does not change, i.e. log S(t) is a linear function of t, indicates a hazard rate function which is constant within the interval. This device of plotting survival functions on a logarithmic scale is especially appropriate for comparing two or more groups. If the curves thus obtained run parallel on some interval on the time axis, one can conclude that the failure rates within that interval do not differ between the groups. In figure 1 this is demonstrated for two arbitrary populations by showing hypothetical failure rates and their corresponding survival functions. It is clear from the figure that an increasing or decreasing failure rate can be more easily recognized in the logarithmically transformed survival function as compared to its linear representation.
Another more fundamental advantage in using the failure rate function $\lambda(t)$ in describing the pattern of events along time holds when the population of interest exists only a limited period of time, with the length of the period possibly differing from individual to individual. As an example one may consider a study of the occurrence of the Sleep Deprivation Syndrome (Schwab, 1994) in patients admitted to an intensive care unit. At the time of discharge from the unit, the patient leaves the population and can not be considered to be at risk anymore. Other examples are

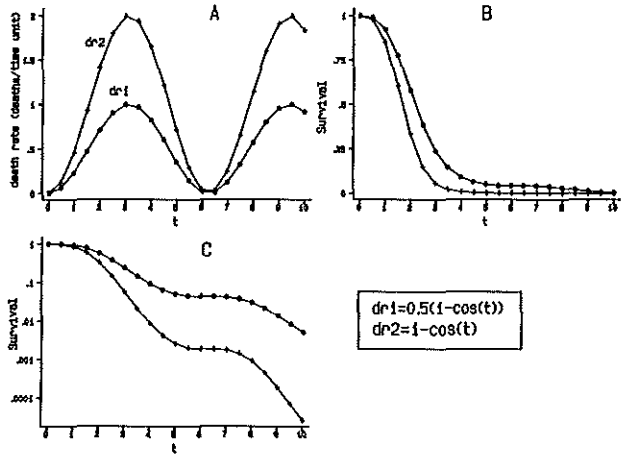


Figure 1. Two hypothetical failure rates (panel A) and their corresponding survivor functions (panel B: linear vertical axis, panel C: logarithmically transformed vertical axis).

studies of the occurrence of some viral infection during dialysis while patients are waiting for a kidney transplant or patients getting bedsores as long as they are bedridden. In these cases patients can be considered to leave the population at the time when the high-risk period ceases. In these circumstances the failure rate function has a clear interpretation and discribes the probability function of failure at time t among those patients still belonging to the population of interest at that time. No such clear interpretation is possible for the function $S(t)=\exp(-\Lambda(t))$. It should especially be noted that its complement 1-S(t) under these circumstances can not be considered to quantify the risk of the event to occur before time t while belonging to the population. If the latter probability is needed, it has to be calculated according to the formula:

$$\int_0^t \lambda(\tau)G(\tau)d\tau \qquad (1.7)$$

Here $G(\tau)$ denotes the probability of being at risk for the event at time τ, i.e. the probability of population membership at time τ without having experienced the event of interest before. Equation (1.7) is used in chapter 5 to assess the lifetime risk of recurrent dysphagia in patients operated for oesophageal carcinoma.

2. Parametric survival functions

In the previous section survival distributions were considered in a general framework. Practical applications may require selection of a particular statistical model. Some often used survival functions are shown in table 1. Each of them is characterized by one or more parameters defining the survival distribution within its class.

Table 1. Some well-known survival functions S(t) and corresponding hazard rate functions λ(t).

Distribution	S(t)	λ(t)
Negative exponential	$\exp(-\alpha t)$	α
Weibull	$\exp(-(\alpha t)^{\beta})$	$\alpha\beta(\alpha t)^{\beta-1}$
Gompertz	$\exp((\alpha/\beta)(1-e^{\beta t}))$	$\alpha e^{\beta t}$
Rayleigh	$\exp(-\alpha t-\frac{1}{2}\beta t^2)$	$\alpha+\beta t$
Log-logistic	$1/(1+(t\alpha)^{\beta})$	$\beta t^{\beta-1}\alpha^{\beta}/(1+(t\alpha)^{\beta})$

Other important distributions are the lognormal distribution and the gamma-distribution. Both these distributions are not included in the table because they do not have a closed form expression.

The negative exponential distribution, which is characterized by a constant hazard rate, was used by Zelen (1966) in a study of survival times in an animal leukemia model. A remarkable feature of the negative exponential failure time distribution is its "lack of memory". If this distribution applies, the future lifetime of an individual does not depend on the lifetime already achieved. If the constant hazard rate is denoted by α, the mean and median lifetime are $1/\alpha$ and $\log(2)/\alpha$, respectively . The negative exponential function S(t) reduces to a straight line when plotted on a logarithmic scale. Metcalf (1974) has shown by extensive examples that for patients with various types of cancer, when survival is studied in patient groups that are relatively homogeneous with regard to prognosis, the survival data can often fairly well be described by the negative exponential distribution.

The Weibull distribution is determined by two parameters, α and β. Because β determines the shape of the hazard function, it is referred to as the shape parameter. The parameter α generally is called the scale parameter. The hazard rate increases with time when the shape parameter β is greater than 1, and decreases when β is less than 1. The Weibull distribution reduces to the negative exponentional distribution in case β equals one. The Weibull model was used by Peto et al (1972) to analyze carcinogenesis experiments.

The Gompertz distribution is often used to model mortality from natural causes in old age. Albertsen et al (1995) used this distribution to model expected survival in prostate cancer patients. The Rayleigh distribution, also called the linear exponential distribution, is an extension of the negative exponential distribution. It is characterized by a linearly increasing or decreasing hazard rate. The model was used by Carbone et al (1967) in the study of survival of myeloma patients.

The log-logistic distribution is the only distribution of those given in Table 1 whose hazard function does not necessarily monotonically increase or decrease along time. The hazard function $\lambda(t)$ has a single maximum for values of β greater than 1, while it is monotonically decreasing along time if $\beta < 1$. The log-logistic distribution was used by Pocock et al (1982) to model mortality for patients with breast cancer.

The lognormal distribution applies when the logarithm of the survival time follows a normal distribution. The associated hazard function is initially increasing with time to a maximum, and is decreasing thereafter. Feinleib and MacMahon (1960) used the lognormal distribution for the discription of survival of patients with chronic lymphocytic leukemia.

Cox and Oakes (1984) showed that the negative exponentional, Weibull, gamma, lognormal and log-logistic distributions can all be put in the framework of the single parametric family of the generalized F-distribution. This property may be of use in discriminating between various candidate distributions in applications.
Other less used parametric survival distributions are discussed by Gross and Clark (1975) and Lee (1980).
All the failure time models described above are of the "proper" type, i.e. the survival function approaches zero for increasing values of t. This means that all members of the population ultimately will experience the event of interest. However, in many cases this will not apply. The distributions described so far may be useful in such applications if not the whole time-axis is to be considered, but only the first part of it. Alternatively, the models can be adapted to situations in which only a fraction of the population is due to experience the event ultimately. This leads to a characterization of the population through a mixture of distributions. An (unspecified) fraction of the population will not experience the event of interest, while for the remaining group a particular distribution of times until the event is assumed to apply. Boag (1949) used this approach in investigating cure rates of cancer patients using a lognormal distribution of survival times for non-cured patients. Farewell et al (1977) estimated the lifetime risk of developing breast cancer and assumed that in those patients developing the disease the age distribution followed a Weibull distribution.
If a population can be considered to be a mixture of populations, while each subpopulation can be characterized by its own hazard function, the hazard function for the population as a whole can be obtained. For instance, if a population is a mixed population with a fraction p of cases having a constant hazard rate λ_1. If the remaining subpopulation has a constant hazard rate λ_2, then it is easy to show that the resulting hazard rate for the whole population can be written as

$$(p\lambda_1\exp(-\lambda_1 t)+(1-p)\lambda_2\exp(-\lambda_2 t))/(p\ \ \exp(-\lambda_1 t)+(1-p)\exp(-\lambda_2 t)).$$

This function will be a decreasing function with advancing time t, and after sufficient time the function will approach a constant level. This level will equal the smallest of the two values λ_1 and λ_2. This is shown in figure 2 where the hazard rate function is shown for a mixed population for some arbitrary chosen parameters p, λ_1 and λ_2. The decreasing hazard rate for the mixture can be explained by selection: the cases with the highest hazard rate will leave the population at a high rate, and ultimately only the group remains with the lowest hazard rate. This example demonstrates that an observed decreasing failure rate for a population is not necessarily due to an intrapatient decreasing activity along time of the biological mechanism which leads to failure. It may even be that each individual's relapse rate increases along time, but that the heterogeneity among individuals is great enough so that the relapse rate of the population as a whole appears to decrease. If one would inappropriately assume that the population is homogeneous in such instances, conclusions may be obtained which are wrong quantitatively as well as qualitatively. If interest focusses on the behaviour of the hazard rate along time, it is imperative to take account of prognostic factors which determine any heterogeneities.

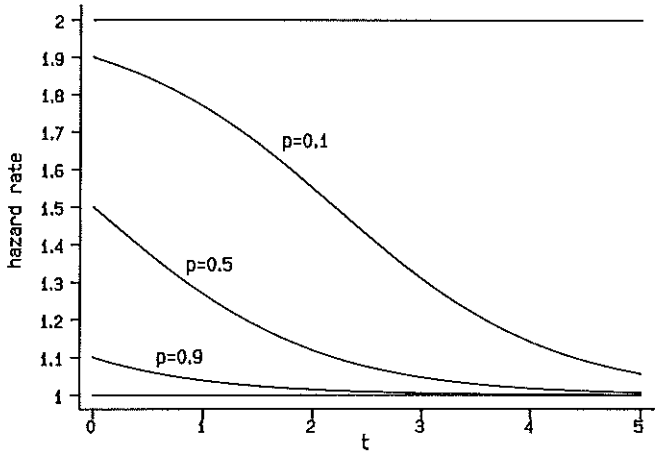


Figure 2. Hazard rate function for a mixed population. The fraction of cases with hazard rate $\lambda_1=1$ is denoted by p; the hazard rate of the remaining subpopulation equals $\lambda_2=2$.

3. Estimation of parametric survival distributions

Due to time limitations or for other reasons in clinical studies, the full time period until the event of interest will often not be available for some patients at the time of analysis. For such patients the event did not occur during the period they were on study, and the actual failure time is (right) censored. Such censored event times occur particularly in clinical trials in which analysis starts relatively soon after entry into the study of the last patient.

Another source of such incomplete follow-up data is caused by patients who drop out of the study after varying periods of time and who therefore are no longer available for further follow-up. It may also happen that some patients still participate in the follow-up regarding clinical aspects, but are no longer available, possibly due to refusal, to perform the tests which are necessary to assess whether the event of interest occurs. It is obvious that such sources of incomplete observations are potentially hazardous with regard to correct interpretation of results. Their presence may lead to biased results in case the mechanism of censoring is in some way related to the probability of experiencing the event thereafter. To perform a meaningful survival analysis in case of such incomplete data, it is required that the censoring mechanism is non-informative, meaning that at all times t the patients still under observation can be regarded as a representative sample of the patients in the population about which inferences are to be made.

If this condition applies, the likelihood L of the sample with n_1 failures and n_2 censored lifetimes with total sample size $n=n_1+n_2$, can be written as the product of the probability density functions for patients having experienced the event, and the survival probabilities for patients whose lifetimes were censored. If δ_i is an indicator which equals 1 in case the lifetime t_i for the i-th patient was actually observed, and equals 0 in case that lifetime was censored, the likelihood L becomes

$$L=\prod_{i=1}^{n} f(t_i)^{\delta_i}\, S(t_i)^{1-\delta_i} \tag{3.1}$$

which can also be written as

$$L=\prod_{u} f(t_i)\prod_{c} S(t_j) \tag{3.2}$$

Here the products Π_u and Π_c denote the products over the uncensored and censored lifetimes respectively. Another way of writing the likelihood is

$$L=\prod_{u} \lambda(t_i)\prod_{j=1}^{n} S(t_j) \tag{3.3}$$

in which the second product concerns all patients in the study, irrespective of whether they had died or not. If the parametric distribution of lifetimes is characterized by a vector ß consisting of p elements whose values are to be estimated, an estimate $\hat{ß}$ can be obtained by maximizing the likelihood function L. Equivalently, the maximum of the logarithm of the likelihood function can be determined. The latter function can be written as

$$\log\, L(\beta)=\sum_{u} \log\, \lambda(t_i,\beta)+\sum_{i=1}^{n} \log\, S(t_i,\beta) \tag{3.4}$$

Finding $\hat{ß}$ such that $L(\hat{ß})$ reaches its maximal value is equivalent to finding the solution $\hat{ß}$ of the p likelihood equations

$$\frac{\partial \log L(\beta)}{\partial \beta}=0 \tag{3.5}$$

Unfortunately, this equation can generally not be solved analytically. An explicit estimate of ß is not available in those cases and an iterative procedure is required to estimate the beta-parameters. The mostly used iterative method is the Newton-Raphson procedure: starting from an initial guess $\hat{ß}^{o}$, after each cycle the updated estimate of ß is obtained from:

$$\hat{\beta}^{j+1}=\hat{\beta}^{j}-\left[\frac{\partial^2 \log L(\hat{\beta}^{j})}{\partial \beta^2}\right]^{-1} \frac{\partial \log L(\hat{\beta}^{j})}{\partial \beta} \tag{3.6}$$

Iteration stops when the difference between the estimate of ß and its updated version differs less than some prespecified limit. Using the Newton-Raphson method one also obtains an estimate of the covariance matrix of the estimate $\hat{ß}$ from the inverse of the sample information matrix, which is given by

$$-[\frac{\partial^2 \log L(\hat{\beta})}{\partial \hat{\beta}^2}] \qquad (3.7)$$

The estimator of ß can be treated as approximately multivariately normally distributed with mean value ß and covariance matrix equalling the inverse of (3.7). Standard errors of the components of $\hat{ß}$ are obtained by taking the square root of the diagonal elements of the estimated covariance matrix.
In case the lifetimes follow a negative exponential distribution with the constant hazard rate λ, an analytic solution for $\hat{ß}$ is obtained. If one reparametrizes λ to $e^{ß}$, the loglikelihood function (3.4) can simply be written as

$$\log L(\beta) = \beta d - e^{\beta} \Sigma t_i \qquad (3.8)$$

in which d represents the number of observed failures. Substitution in the likelihood equation (3.5) leads to the estimate of the hazard rate $e^{\hat{ß}} = d/\Sigma t_i$. From (3.7) it then follows that the standard error of $\hat{ß}$ equals $1/\sqrt{d}$. Two-sided $(1-\alpha)$-confidence limits for λ are obtained from

$$e^{\hat{\beta} \pm Z_{1-\alpha} / \sqrt{d}} \qquad (3.9)$$

Here Z_p denotes the $100p^{th}$ percentile from the standard normal distribution.
The reparametrisation used has the advantage that the estimator of ß approaches a normal distribution more rapidly than the estimator of λ itself due to the restriction to positive values for λ.
The confidence limits (3.9) are approximate and are derived from asymptotic normality properties. Exact confidence limits are available in cases where there is no censoring or censoring is of the kind which is so-called Type II [Miller, 1981]. Type II censoring occurs for instance in experiments with animals which are put on test simultaneously and testing is stopped after a prespecified number of animals have developed the event of interest.
For purposes of illustration of the foregoing, survival data of 174 patients operated for kidney cancer are used here (these data are more fully discussed in chapter 3). Assuming an exponential survival distribution, the (constant) death rate was estimated to be 0.011 deaths/month, which value corresponds to a mean survival time of 90 months. It appeared however that the assumed negative exponential model did not fit the survival data very well. In comparing the estimated survival curve arising from the assumed negative exponential model with the observed survivor curve which can be obtained without making any assumptions about the parametric form (to be discussed in section 5 below), both curves do not seem to agree very well. This is shown in figure 3.

In fitting the Weibull model to the data, it appeared that the hazard rate decreased along time as shown by the shape parameter $ß = 0.75$, which is significantly smaller than 1 ($p < 0.01$; likelihood-ratio test). However, this finding does not imply that the risk of dying for individual patients decreases along time. As will be shown in chapter 3, the patient group appears to be rather heterogeneous with respect to various

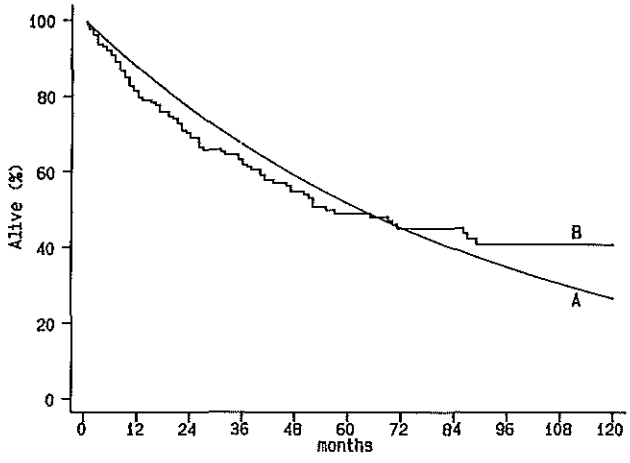


Figure 3. Estimed survival curve assuming a negative exponential survival function (curve A) and non-parametric estimate (Kaplan-Meier estimate: curve B).

important prognostic factors. For patients with a similar prognosis it is demonstrated in section 8 that the hazard rate is rather constant along time. The decreasing hazard rate for the group as a whole should be explained by the selection mechanism already described. High risk patients die relatively soon whereas patients with lower risk remain in the cohort.

4. Testing a known distribution against sample data.

Sometimes there is knowledge about a specific survival distribution $S_0(t)$, possibly derived from the litterature, and one may need to test obtained survival data against that specified survival distribution. A test which focuses on the maximal distance between the non-parametric survival curve and the postulated curve, is available [Koziol, 1980]. A more powerful test is obtained if testing occurs against special alternatives. If it is assumed that the hazard function describing the data can be written as $\lambda(t)=e^{\beta}\lambda_0(t)$, where the hazard rate function $\lambda_0(t)$ corresponds to the survivor function $S_0(t)$, then it follows from equation (1.6) that S(t) follows the so-called Lehmann-alternative: $S(t)=S_0(t)^{e^{\beta}}$. [Lehmann, 1953] The loglikelihood (3.4) can then be written as

$$\log L(\beta)=d\beta-e^{\beta}\sum\Lambda_o(t_i)+\sum\log\lambda_o(t_i) \qquad (4.1)$$

Here Σ denotes the sum over all patients in the study. Maximizing this likelihood gives $e^{\beta}=d/\Sigma\Lambda_o(t_i)$. The score test for testing $\beta=0$ is based on the ratio of the first and the negative of the second derivative of the loglikelihood function, both evaluated at $\beta=0$ [Cox and Hinkley, 1974]. This leads to the so-called one-sample logrank test statistic [Harrington and Fleming, 1980]:

$$T = \frac{d - \Sigma\Lambda_o(t_i)}{\sqrt{\Sigma\Lambda_o(t_i)}} \qquad (4.2)$$

which follows approximately the standard normal distribution.
If $\lambda_o(t)$ is allowed to differ between individuals, for instance in comparing observed survival of a patient group with expected survival according to vital statistics of a group matched for age and sex from the national population, then the estimate of e^{β} represents the standardized mortality rate [Breslow and Day, 1987].

5. Nonparametric estimates of survival distribution

In cases where there is insufficient information about the parametric form of the survival distribution, the estimator of S(t) which is generally used, is the Kaplan-Meier estimator. This estimator is also called the product-limit estimator [Kaplan and Meier, 1958]. If among the n patients studied there are k different observed ordered failure times $t_1 < t_2 < \ldots < t_k$ and at each timepoint t_i (i=i,...,k)there are r_i patients still at risk (i.e. they all have an observation period of at least t_i), of whom the number of patients who died is given by m_i, the Kaplan-Meier estimate of the proportion surviving t is given by

$$\hat{S}(t) = \prod_{t_i \le t} \frac{r_i - m_i}{m_i} \qquad (5.1)$$

Here the product runs over all observed failure times t_i up to t. If the largest observation time corresponds to a censored value, the estimator of the survival function is not defined beyond that time. The estimate is a step function with constant levels between the observed failure times and jumps at the recorded failure times. The Kaplan-Meier estimator is asymptotically unbiased and approaches a normal distribution [Breslow and Crowley, 1974]. The estimated variance of S(t) is given by

$$\hat{S}^2(t) \sum_{t_i \le t} \frac{m_i}{m_i(r_i - m_i)} \qquad (5.2)$$

and is commonly referred to as Greenwood's formula (1929).
For the purpose of confidence interval construction it is to be preferred [Kalbfleisch and Prentice, 1980] to base that interval on the asymptotic variance of $\log(-\log(\hat{S}(t))$, which reads

$$\{\sum_{t \le t_i} \frac{m_i}{m_i(r_i - m_i)}\} \{\sum_{t \le t_i} \log \frac{r_i - m_i}{r_i}\} \qquad (5.3)$$

Apart from the improved adequacy of the normal approximation, using this variance estimate will avoid impossible values for the confidence limits outside the range [0,1]. Alternative confidence limits have been proposed by Thomas and Grunkemeyer (1975).

Brookmeyer and Crowley (1982) have described a very simple graphical device to determine confidence limits for the median survival time. The lower limit corresponds to the smallest value of t at which the confidence interval for S(t) includes the 50% survival probability. The upper limit corresponds to the smallest value of t greater than the median for which the confidence interval for S(t) fails to include the 50% probability. Gill (1980) has developed confidence bands for the whole survival function on some fixed interval [0,t].
An alternative estimator of the survival function is Nelson's estimator (Nelson, 1969). This estimator is given by

$$\hat{S}(t) = \exp\{-\hat{\Lambda}(t)\} \tag{5.4}$$

in which the estimator of the cumulative hazard rate is given by the step function

$$\hat{\Lambda}(t) = \sum_{t_i \le t} \frac{m_i}{r_i} \tag{5.5}$$

There will be no large differences between the Kaplan-Meier estimate and Nelson's estimate if the number of deaths at each time point is relatively small in comparison with the numbers at risk. The minor difference between the two estimators of survival for the kidney cancer survival data already discussed, is shown in figure 4. The solid line denotes the Kaplan-Meier curve, while the dotted line is corresponding to Nelson's estimate. Both estimates are hardly distinghuisable; they differ at most 0.4%.

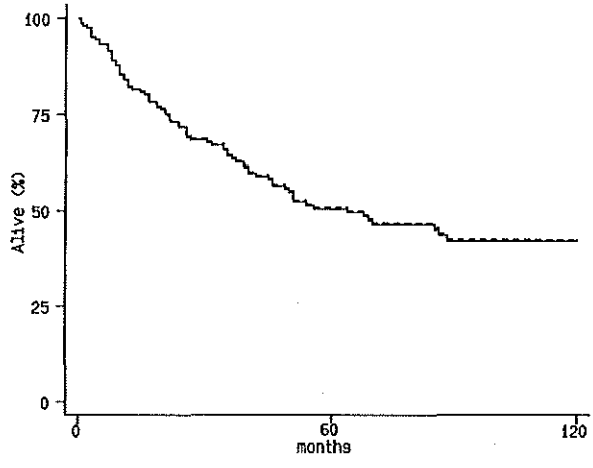


Figure 4. Kaplan-Meier (solid curve) and Nelson (dotted curve) estimates of survival for operated renal cell carcinoma patients.

6. Comparisons of survival curves

In comparing failure time data between two or more groups, various powerful distribution-free test statistics are available. A test statistic which considers the

maximum difference between two survival curves has been described [Fleming et al, 1980]. The two tests mostly used, however, are the logrank test and a generalized Wilcoxon test.
In the following the logrank test and a generalized Wilcoxon test are described for the comparison of two groups. Extension to a comparison of more than two groups, including a test for trend in case of ordered alternatives, is straightforward [Thomas et al, 1977].
If in the two groups (A and B) combined there are k different observed failure times, these can be denoted by $t_1 < t_2 < \ldots < t_k$.
m_{iA} denotes the number of events observed, and r_{iA} denotes the number of patients at risk at time t_i (i=1,...,k)in group A. Further, m_i and r_i (i=1,...,k)denote the respective numbers in both groups combined. The logrank test and the generalized Wilcoxon test can both be written as

$$T=\sum_{i=1}^{k} w_i(m_{iA}-m_i r_{iA}/r_i) \tag{6.1}$$

T represents a weighted sum of the differences between the observed number of deaths in sample A and the conditionally expected number under the hypothesis of equal survivorship.
If this hypothesis holds, the variance of T is a weighted sum of hypergeometric variances, which can be written as

$$V=\sum_{i=1}^{k} w_i^2 \frac{r_{iA} r_{iB} m_i (r_i-m_i)}{r_i^2 (r_i-1)} \tag{6.2}$$

In this formula r_{iB} denotes the number of patients at risk for the event at time point t_i in group B.
The test statistic $T/V^{1/2}$ follows approximately a standard normal distribution.
By taking weights $w_i=1$ for all i, the logrank test is obtained. If the weights are taken to be

$$w_i=\prod_{j=1}^{i}\{(r_j-m_j+1)/(r_j+1)\} \tag{6.3}$$

a generalized Wilcoxon test is obtained [Prentice and Marek, 1979]. The weights (6.3) correspond closely to the value of the single Kaplan-Meier survival curve at t_i when both groups are combined.
The logrank test is demonstrated [Mantel, 1966] to be an adaptation of the Mantel-Haenszel procedure [Mantel and Haenszel, 1959] for censored survival data. The logrank test is most powerful in cases where the hazard rate functions in the various groups are proportional to each other [Peto, 1972]. When there are no censored lifetimes, the generalized Wilcoxon test reduces to the well-known Wilcoxon rank-sum test. Small-sample properties of the logrank-test and generalized Wilcoxon test have been investigated by Keller and Chmelevsky (1983).

7. Parametric survival time regression models

For a heterogeneous sample of patients, of which the heterogeneity can be expressed by a number p of covariates $z_1, z_2, \ldots, z_p$, it is needed to assess the relation between these covariates and the time to the occurrence of the event of interest.
One of the earliest models used, although only for uncensored survival data and one covariate z_1, was the negative exponential model [Feigl and Zelen, 1965]. The relation between the expected survival time and the covariate was expressed as a linear model:

$$E(T) = \theta_0 + \beta_1 z_1 \tag{7.1}$$

Using the method of maximum likelihood the parameters θ_0 and β_1 can be estimated from the data. Zippin and Armitage (1966) described procedures to follow for this linear model in case censored event times were present. The multiple linear model, i.e. allowing for more than one covariate in (7.1), was discussed by Mantel and Myers (1971). Alternative models for the univariate case suggested by Feigl and Zelen (1969) were

$$E(T) = \frac{1}{\theta_0 + \beta_1 z_1} \tag{7.2}$$

and

$$E(T) = \theta_0 e^{\beta_1 z_1} \tag{7.3}$$

Models (7.2) and (7.3) can also be interpreted as models in which the hazard rate is associated with z_1 in terms of linear and loglinear functions of z_1. Both models (7.1) and (7.2) suffer from the disadvantage that, unless precautions are taken in the procedure when estimating the parameters, non-permissible negative estimated values for the mean lifetime may occur at extreme values of z_1. This can not happen with model (7.3).
The multiple version of model (7.1), allowing for various covariates, was used to evaluate survival in prostatic cancer patients [Byar et al, 1974]. As already noted, model (7.3) allowing for multiple covariates, can be written as:

$$\lambda(t;z) = \theta_0 e^{\beta z} \tag{7.4}$$

In this formula z denotes a column vector of p covariates and ß denotes a row vector of p associated regression coefficients. This model was used by Prentice (1973) in evaluating survival of lungcancer patients. Prentice (1973) also noted that a greater flexibility could be achieved if the constant term θ_0 in model (7.4) was allowed to depend on time. This leads to the Weibull model

$$\lambda(t;z) = \theta_0 t^{\theta_1} e^{\beta z} \tag{7.5}$$

Other functions of time for the leading term in model (7.5), for instance a polynomial of a certain degree [Taulbee, 1979], may be more appropriate.

An important property which models (7.4) and (7.5) share is that the covariates act multiplicatively on some "baseline" hazard function. This baseline hazard function corresponds to the hazard function of the patients whose covariate values all take the value zero.
Another model which may be appropriate in applications is the linear lognormal model in which the expected value of the logarithm of the survival time is a linear function of the components of the covariate vector z:

$$E(\log T) = \theta_0 + \beta z \tag{7.6}$$

Random deviations of log(T) from the expected value follow a normal distribution with mean zero and standard deviation θ_1.
Models (7.5) and (7.6) belong to a larger class of so-called accelerated failure time models [Kalbfleish and Prentice, 1980]. These can be characterized by the property that the covariates have a multiplicative effect on T in stead of a multiplicative effect on the hazard rate. If an independent censoring mechanism applies, the regression coefficients can be estimated by maximizing the likelihood of the observations. It is allowed that the censoring mechanism depends on z. What is required is that, conditionally on z, the censoring is unrelated to the future risk of the event.
If the sample size equals n and t_i (i=1,...,n) denotes the individual observation times and z_i denotes the individual covariate vectors, the likelihood can be written as

$$L(\beta) = \prod_{i=1}^{n} \lambda(t_i;z_i,\beta)^{\delta_i} \, S(t_i;z_i,\beta) \tag{7.7}$$

Here δ_i indicates that t_i represents a complete failure time ($\delta_i=1$) or a censored one ($\delta_i=0$). The vector ß is taken to represent not only the regression coefficients associated with the various covariates, but the vector includes also other parameters indicating the chosen parametrisation such as θ_0 and θ_1 in model (7.5).
Estimates of ß, as described before for the case without covariates, can be obtained by solution of the likelihood equations using the Newton-Raphson procedure. Variances and covariances can be obtained by using the inverse of the observed matrix of second partial derivatives of the loglikelihood function. These can be used for the construction of confidence limits for individual parameters or for functions thereof. For testing hypotheses about subsets of parameters in a multiple model, which may be useful in reducing the dimensionality of the model by eliminating variables which are of lesser importance considering the presence of other variables in the model, three commonly used tests are available: Wald's test, Rao's score test and the likelihood ratio test [Cox and Hinkley, 1974].
If ß is written as ($ß_A$, $ß_B$), in which $ß_B$ denotes the vector of the subset of parameters for which the hypothesis $ß_B=0$ applies, the Wald test is based on the estimate of $ß_B$ of the full model including $ß_A$ and its estimated covariance matrix. For a single parameter Wald's test reduces to the ratio of the square of the estimate of $ß_B$ and the square of the standard error of the estimate of $ß_B$.
The score test is based on the value of the first derivative of the loglikelihood function and the observed information matrix, both evaluated at $ß_B=0$.
The likelihood ratio test evaluates the difference between the loglikelihood of the full model and the restricted model, i.e. the model with $ß_B=0$.

If the dimension of β_B equals v, and the hypothesis $\beta_B=0$ is true, all three test statistics follow approximately a Chi-square distribution with number of degrees of freedom equalling v. All three test statistics are asymptotically equivalent. There is general agreement, however, that in moderately sized studies the likelihood-ratio test is to be preferred.

Goodness of fit of a chosen parametric model can be graphically evaluated by using generalized residuals [Cox and Snell, 1969]. If $\hat{S}_i(t_i,z_i,\hat{\beta})$ denotes the estimated survival function based on the model for the i^{th} patient evaluated at t_i, then, if the model is correct, the values $\hat{S}_i(t_i,z_i,\hat{\beta})$ should approximately behave as an uniform distribution on the interval [0,1]. Therefore, the residuals -log $\hat{S}_i(t_i,z_i,\hat{\beta})$, which equal the estimated cumulative hazards $\hat{H}_i(t_i,z_i,\hat{\beta})$, should behave approximately as a sample from the negative exponential distribution with mean one. The residuals corresponding to censored survival times are themselves censored residuals. By calculating the Kaplan-Meier curve of the residuals and plotting the curve on a logarithmic scale versus the residuals, approximately a straight line must emerge with a slope equal to -1 (an example is shown in figure 7 of the next section). In case of continuous covariates, the residuals can be plotted against the covariates. Here the expected remaining value of 1 for the standard negative exponential distribution can be added to the residual in case the residual is a censored one.

The evaluation of adequacy of fit for a particular model as a special case of a larger, more general, family of models is described by Farewell and Prentice (1977).

8. Cox regression

Since the use of hazard rates is basic in many applications and conceptually simple, the modeling of the hazard rate $\lambda(t;z)$ as a function of time and covariates seems attractive. Various parametric models already discussed were of this type. A disadvantage of these parametric models, however, is that the functional relation with time has to be specified in detail up to a few parameters which may be estimated.

This disadvantage had led Cox (1972) to formulate the model:

$$\lambda(t;z,\beta)=\lambda_0(t)e^{z\beta} \qquad (8.1)$$

The remarkable characteristic of this model is that $\lambda_0(t)$ is completely left unspecified. The model (8.1) is the one mostly used in practice. The term $\exp(z\beta)$ in the model, however, may occasionally need to be replaced by another function of z and β. Generally the model is referred to as the proportional hazards model, because the covariates affect the so-called reference hazard rate function $\lambda_0(t)$ in a multiplicative manner.

The fact that there is no need to specify the reference function $\lambda_0(t)$ gives the model an extreme flexibility. With the introduction of partial likelihood [Cox, 1975] it became possible to estimate the p elements of the column vector of unknown regression coefficients β associated with the covariate vector $z=(z_1,z_2,\ldots,z_p)$The partial likelihood function can be used in the same manner as ordinary likelihoods for

asymptotic inference on ß. The method also allows for the estimation of the reference hazard rate function $\lambda_0(t)$. This function, also known as the baseline hazard rate function, can be thought to apply if all covariates take the value of zero. Using survival functions, the model can equivalently be expressed as

$$S(t;z) = S_0(t)^{e^{z\beta}} \tag{8.2}$$

where $S_0(t)$ denotes the reference survival function corresponding to the covariate vector $\beta=0$.
Suppose there are K patients who experienced the event, and these patients all had different event times $t_1 < t_2 < \ldots < t_K$; for each $t_i (i=1,\ldots,K)$ the set of patients still under observation at time t_i will be represented by $\mathbf{R}(t_i)$. The partial likelihood can then be written as

$$L(\beta) = \prod_{i=1}^{K} \{e^{z_{(i)}\beta} / \sum_{j \in R(t_i)} e^{z_j\beta}\} \tag{8.3}$$

The product runs over the K event times and the term $z_{(i)}$ in the numerator denotes the covariate vector of the patient who experienced the event at time t_i. Several generalisations of (8.3) have been proposed to accomodate tied failure times [Kalbfleisch and Prentice, 1980]. These resulting likelihoods, however, are rather complicated. In practice therefore use is mostly made of an approximate likelihood [Breslow, 1974]:

$$L = \prod_{i=1}^{K} \{e^{S_{(i)}\beta} / (\sum_{j \in R(t_i)} e^{z_j\beta})^{m_i}\} \tag{8.4}$$

In this formula m_i $(i=1,\ldots,k)$ denotes the number of patients who died at time t_i, while $S_{(i)}$ $(i=1,\ldots,k)$ denotes the sum of the covariate values for these m_i individuals.
This approximate likelihood will suffice in case the fraction of tied failure times is not too large. If there are no tied failure times, likelihood (8.4) reduces to likelihood (8.3). Maximizing the likelihood leads to the estimates of the elements of the unknown parameter vector ß. The available likelihood procedures give rise to straightforward significance test about elements of ß and the calculation of approximate confidence limits [Tsiatis, 1981]. Using simulation, small sample properties were examined by Johnson et al (1982). It is worth mentioning that, in case z is a covariate vector denoting group membership, the score test of the hypothesis $\beta=0$, leads to the well-known logrank test.
To estimate $\lambda_0(t)$, Breslow (1974) has proposed a step function which has a constant value between the different failure times. In the interval $(t_{i-1}, t_i]$, with t_0 taken to be zero, his estimator takes the form:

$$\hat{\lambda}_0(t) = \frac{1}{(t_i - t_{i-1})} \frac{m_i}{\sum_{j \in R(t_i)} e^{z_j \beta}} \tag{8.5}$$

Smooth estimates of $\lambda_0(t)$ are described by Whitmore and Keller (1986) and Ramlau-Hansen (1983). The cumulative reference hazard function $\Lambda_0(t) = \int_0^t \lambda_0(t)dt$ may be estimated by

$$\hat{\Lambda}_0(t) = \sum_{t_i \le t} m_i / \sum_{j \in R(t_i)} e^{z_j \hat{\beta}} \tag{8.6}$$

A slightly different estimate of the cumulative reference hazard function is proposed by Link (1984). The estimate of the baseline survival function $S_0(t)$ can be obtained by exponentiating the negative of $\hat{\Lambda}_0(t)$. If all elements of ß are zero, i.e. the covariates considered have no prognostic value, the cumulative hazard rate estimator reduces to the step function:

$$\hat{\Lambda}_0(t) = \sum_{t_i \le t} \frac{m_i}{r_i} \tag{8.7}$$

where r_i (i=1,...,k)denotes the number of patients at risk at t_i. The corresponding survival function estimator equals the Nelson estimator (equation 5.4) in this case.
For illustration, the data of the renal carcinoma patients (chapter 3) are used here. Among various factors evaluated, the most important factors which were found to be related to survival after operation were the gender of patients (z_1=1 if male, z_1=0 if female) whether or not there was renal vein invasion at pathological examination (z_2=1 if yes, z_2=0 if no) and the level of the erythrocyte sedimentation rate (z_3=ESR (mm/hr)). The resulting prognostic model could be written as

$$\lambda(t;z_1,z_2,z_3) = \lambda_0(t) e^{PI(z_1,z_2,z_3)} \tag{8.8}$$

in which the prognostic index PI equals $0.79\, z_1 + 0.89\, z_2 + 0.018\, z_3$. Larger values of PI indicate a worse prognosis. The estimate of the reference hazard rate function $\lambda_0(t)$ based on this model is shown in figure 5. Apart from some random variation, the estimated reference function does not seem to vary systematically along time. Fitting the model again, but now with $\lambda_0(t)$ taken to be a constant (λ_0), thus in fact fitting the negative exponentional model (7.4), the resulting estimate of λ_0 equals 0.0017 (95% confidence interval:0.0013-0.0022) deaths/month. Fitting of the negative exponentional model did not lead to appreciable changes with respect to the estimates of the regression coefficients or their standard errors.
The prognostic index alone is not sufficient to assess the impact on survival probabilities for individual patients. In addition to the value of the prognostic index, the estimated reference survival function using equation (8.2) is needed. Figure 6 shows estimated 5- and 10-years survival probabilities according to values of the prognostic index for the renal carcinoma patients.

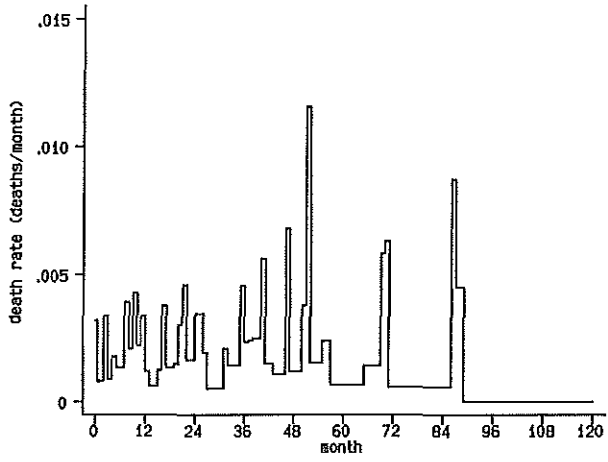


Figure 5. Estimated death rate function for operated renal cell carcinoma patients whose prognostic index PI equals 0.

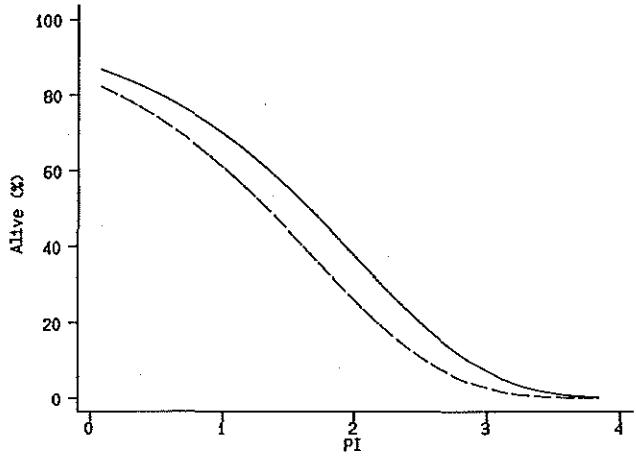


Figure 6. Five years (solid curve) and ten years (dotted curve) survival probabilities according to prognostic index PI.

Confidence limits for the survival probabilities S(t;z) at some fixed time t have been derived by Andersen et al (1983). Confidence bands for the whole curve S(t;z) for some fixed z have been derived by Lin and Fleming (1991). It is not possible to derive confidence limits for curves such as shown in figure 6. This is due to the fact that standard errors of the estimated survival probabilities do not only depend on the value of the prognostic index, but also depend on the precise values of the covariables resulting in a particular value of the prognostic index [Altman and Anderson, 1986].

9. Goodness-of-fit for the proportional hazards model

Tests for goodness-of-fit are described by Schoenfeld (1980) and Andersen (1982). Because of their omnibus character these tests are not very powerful and are therefore seldom used. Better suited are tests which are sensitive to particular aspects thought to be important. One such aspect is the stability of the resulting regression coefficients when the model is fitted to different intervals of the time-axis. By using likelihood ratio tests the significance of differences in outcomes between the various intervals can be assessed easily. Score-tests to be used with such a-priori determined intervals have been described by O'Quigley and Pessione (1989). Any relation between the values of ß and time can also be investigated by introducing additional terms in the model which are the product of one or more components of the covariate vector z and functions of time itself. This way smooth changes along time of the effect of covariates can be assessed [Cox, 1972].
Graphical procedures to evaluate various aspects of fit are also very useful. One such method is to stratify the patient group according to some covariate values. The basic model (equation 8.1 above) can be extended by allowing the reference hazard functions for these strata to differ from each other. If the covariates used in the formation of these strata have a multiplicative effect on a common baseline hazard function, the various estimated hazard functions should be proportional to each other. If this assumption is true, then the various cumulative hazard functions should run parallel on a logarithmic scale.
As was the case with the parametric models discussed before, the assessment of fit by using generalized residuals can also be done for proportional hazard models [Kay, 1977]. If the residual R_i for the i^{th} patient with observation time t_i is defined as -log $S_i(t_i)$, the residual for that patient can be estimated by

$$\hat{R}_i = \hat{\Lambda}_0(t_i) e^{z_i \hat{\beta}} \qquad (9.3)$$

If the model holds, the generalized residuals should follow a negative exponential distribution with mean one. A first step in the evaluation of fit using residuals is to plot the "survivor" function of the residuals with a logarithmically transformed vertical axis. The Kaplan-Meier representation should approximate a straight line then. For the kidney cancer survival data (chapter 3), in figure 7 the Kaplan-Meier curve is plotted of the generalized residuals obtained after fitting the proportional hazards model allowing for gender, renal vein involvement and ESR. Based on this graph the applied model seems to fit well because the curve does not greatly differ from its expectation.
The residuals can also be investigated for groups of patients or plotted against particular covariates. For example, Damhuis and Blom (1995) found in a population based study that survival for renal carcinoma patients decreased with increasing age at diagnosis. No such relation was found, neither univariately nor multivariately, for the patients described in chapter 3 . Figure 8 shows the generalized residuals, after fitting the proportional hazards model taking account of gender, renal vein involvement and ESR, which are plotted against the age of patients at diagnosis. No relation between the residuals and age is apparent, indicating that in our patient group age is of no additional prognostic value during the follow-up period considered within the framework of the proportional hazards model.

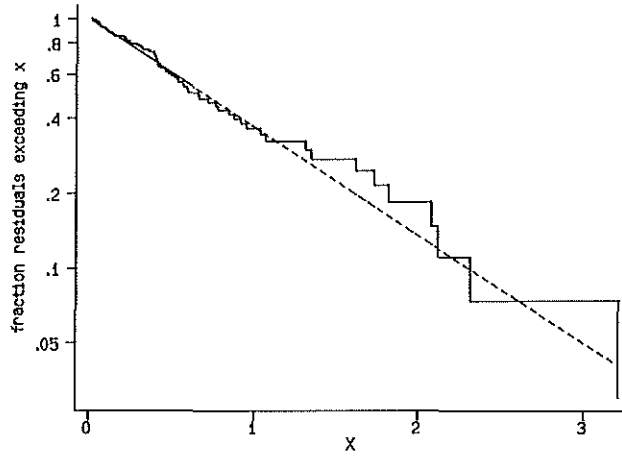


Figure 7. Kaplan-Meier representation of distribution of generalized residuals. Dotted line denotes the standard exponential distribution.

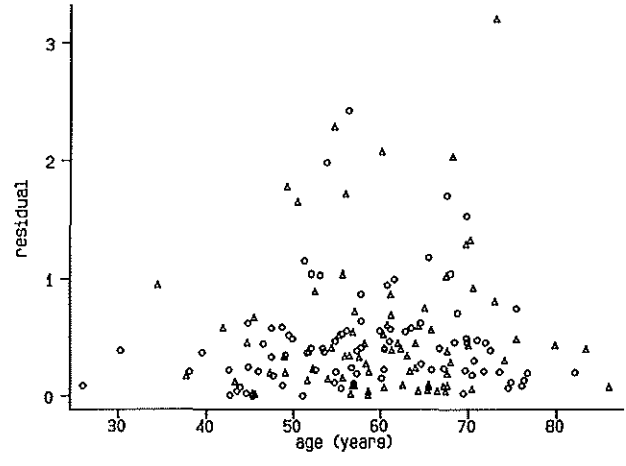


Figure 8. Generalized residuals (circles: patients alive, triangles: dead patients) plotted against the age of patients at diagnosis.

Various other type of residuals have been defined by Therneau et al (1990). Diagnostic plots that are useful for assessing the effect of adding a variable, detecting non-linearity, or identification of influential points, are described by Chen and Wang (1991).

A noteworthy characteristic of the proportional hazards model is that if this model holds for a certain combination of covariates, the model does not apply any more if one or more relevant covariates are discarded from the model. This is true irrespective of the distribution of such omitted covariates. The effect of dropping a

balanced variable in the proportional hazards model is shown by example in the following. Assume that the proportional hazards model $\lambda(t;z_1,z_2)=\lambda_0(t)\exp(0.693z_1 +0.693z_2)$ holds. Here z_1 represents a binary (0,1) variable denoting treatment group, and z_2 denotes a binary (0,1) indicator for some prognostic factor whose associated effect is to double the hazard rate. According to the model, the ratio of hazard rates for the two treatment groups also equals 2, both for $z_2=0$ and $z_2=1$. Taking λ_0 equal to 1, figure 9 is obtained. This figure shows at each point in time the resulting ratio of hazard rates of the two total treatment groups for varying fractions p of cases with $z_2=1$ (p=0.1, 0.5 and 0.9). With increasing prevalence of the factor associated with the greater risk ($z_2=1$), the deviations from a constant hazard ratio of 2 become more pronounced. This example demonstrates that in comparing treatment groups the omission of an important covariate in a situation that the proportional hazard model applies, even when this covariate is balanced between the groups, will lead to a biased estimate of the treatment effect. This effect will then be underestimated, contrasting the situation in ordinary multiple linear regression. A quantitative study of this biasing effect was performed by Chastang et al (1988). The loss of power due to the omission of a balanced covariate was studied by Lagakos and Schoenfeld (1984).

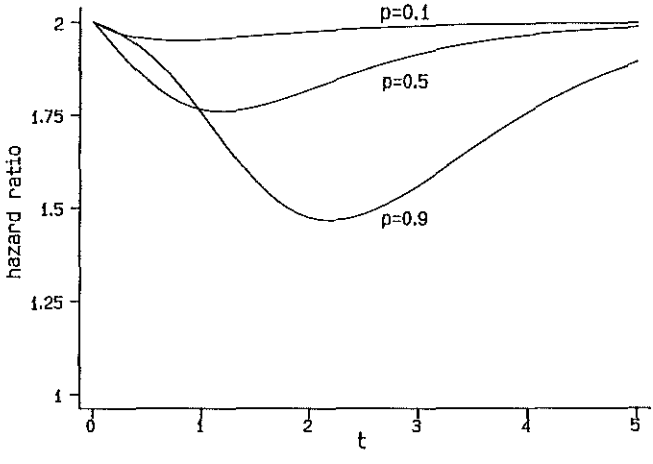


Figure 9. Effect of an omitted balanced covariate in a proportional hazards model. The prevalence of the omitted prognostic factor is denoted by p.

10. Discussion and introduction to chapters 3-6.

With the increased use of controlled clinical trials, there has been a proliferation of statistical methodology for the analysis of survival data. Graphical representations of outcomes, e.g. using Kaplan-Meier curves, are important components of analysis. The regression models described, either parametric or nonparametric, have proven useful in identifying factors which are strongly associated with survival in numerous studies. The parametric models are most useful in studies which have a relatively small

number of patients or events, or when it is needed to extrapolate beyond the range of observed failure time data. An advantage of some of the parametric models may also be that their results are more easy to communicate because of their great similarity with the familiar multiple regression method in which the expected outcome is linearly related to a number of factors. With the possibility of allowing the effects of the baseline covariates to depend on time itself, and the possibility to allow for the inclusion of various baseline hazard functions which are related to different subgroups of patients, the Cox-regression method is an extremely flexible tool in analysing rates of occurrence of some well-defined event. An interesting application in a competing risk framework was made by Kay (1986) who studied various factors regarding their impact on different cause-specific death rates (cardiovascular, cancer or other causes) in patients with prostate cancer. In that study it was found that using Cox-regression it was well possible to describe each of the separate cause-specific death rates, but the proportional hazards assumption failed regarding the overall mortality. This example demonstrates that a judicious choice of a subtyping of events, when appropriate, may be advantageous in order to obtain a better understanding of the data.

The proportional hazards regression model as described applies to continuous times, i.e. the times of the occurrence of the event can be observed (nearly) exactly. This requires an intensive observation of patients. In some instances patients are only regularly seen after relatively long intervals . When at a particular follow-up visit it is evident that the event has occurred, e.g. a recurrence of disease, and it is only possible to conclude that the exact time-point of the recurrence lies between the previous visit and the current one, the method should be adapted [Prentice and Gloeckner, 1978]. A method suited for nonsynchronized follow-up intervals is described by Finkelstein (1986).

Another situation may be that the follow-up proceeds in truly discrete units, e.g. the number of cycles women with fertility problems need treatment in order to achieve pregnancy. For such truly discrete situations, Cox (1972) has proposed an analogue of the proportional hazards model. With this method, the probability of the event studied, conditional on not having observed the event before, is written as a logistic regression equation [Cox, 1970].

Occasionaly it may occur that patients do not all enter the study at time "zero", but some time later. This will happen, for instance, when one studies survival from diagnosis in a particular hospital setting, and one wants to include also those patients who had a diagnosis elsewhere but were referred to the particular hospital a considerable time thereafter. The inclusion of such late entries, without applying specific corrections for their late entry, will severely bias the resulting survival rates. A similar bias will occur when studying survival of patients diagnosed in a certain period, if one also wants to include the prevalent cases in the calculations. It is evident that, without adjustments in the calculations, the proportion of long-term survivors will then be over-estimated [Kurtzke, 1989]. Cnaan and Ryan (1989) have descibed methods which appropriately deal with such late entries.

Similar adaptions are required when another choice of time-scale is indicated. For instance, in epidemiological studies regarding the incidence of a particular disease such as cancer, age will often be a more appropriate time-axis than the length of time since some more or less arbitrary time point which marks the beginning of the study. The time-axis most strongly related to the outcome studied is to be preferred [Breslow and Day, 1987]. In most clinical studies the basic time variable corresponds to the

duration of follow-up after diagnosis or start of treatment. However, in studying an untoward event during pregnancy, for instance, it may be more appropriate to use the gestational age as the time-axis instead of measuring time since the moment that women entered the study and came under observation.
In the following three chapters clinical studies are described in which Cox-regression played an important role in the evaluation of baseline characteristics.
In chapter 3 various baseline characteristics are evaluated regarding prognosis in operated kidney cancer patients. Some aspects of this study are already discussed above. One of the more important prognostic factors in that study appeared to be a very low-cost parameter, namely the erythrocyte sedimentation rate. The importance of ESR as prognostic factor in renal cell carcinoma was recently also found by Roosen et al (1994), Fossa et al (1994) and Lopez Hanninen et al (1996).
Chapter 4 gives an account of a randomized study in colorectal cancer which evaluated the potential benefit regarding prognosis of an autologous blood transfusion program as compared to standard allogeneic transfusions. The main goal of Cox-regression in this application was to provide more precise estimates of transfusion effects.
In chapter 5 various prognostic factors are evaluated for patients surgically treated for esophagus carcinoma. In an appendix to that article it is shown that ordinary Kaplan-Meier curves are not suitable for obtaining absolute risk percentages of recurrent dysphagia.
In the foregoing only methods to evaluate factors established at baseline are discussed. In chapter 6 methods are discussed which allow evaluations of observations on patients made during follow-up, thereby providing means to update prognosis.

References

Albertsen PC, DG Fryback, BE Storer, TF Kolon, J Fine. Long-term survival among men with conservatively treated localized prostate cancer. JAMA 1995; 274: 626-631.

Altman DG, PK Andersen. A note on the uncertainty of survival probability estimated from a Cox regression moel. Biometrika 1986; 73: 722-724.

Andersen PK. Testing goodness-of-fit of Cox's regression and life model. Biometrics 1982; 38: 67-77.

Andersen PK, E Christensen, L Fauerholdt, P Schlichting. Measuring prognosis using the proportional hazards model. Scand J statist 1983; 10: 49-52.

Boag JW. Maximum likelihood estimates of the proportion of patients cured by cancer therapy. J Roy Statist Soc 1949; B11: 15-53.

Breslow N. Covariance analysis of censored survival data. Biometrics 1974; 30: 89-99.

Breslow NE, J Crowley. A large sample study of the life table and product limit estimates under random censorship. Ann Statist 1974; 2: 437-453.

Breslow NE, NE Day. Statistical methods in cancer resarch. UICC, Lyon, 1987.

Brookmeyer R, JJ Crowley. A confidence interval for the median survival time. Biometrics 1982; 38: 29-41.

Byar DR, R Huse, JC Bailar. An exponential model relating censored survival data and concomitant information for prostatic cancer patients. J Nat Cancer Inst 1974; 52: 321-326.

Carbone P, L Kellerhaus, E Gehan. Plasmacytic Meyeloma: a study of the relationship of survival to various clinical manifestations ans anomalous protein type in 112 patients. Am J Medic 1967; 42: 937-948.

Chastang C, D Byar, S Piantadosi. A quantitative study of the bias in estimating the treatment effect caused by omitting a balanced covariate in survival models. Statist Medicine 1988; 7: 143-1255.

Chen CH, PC Wang. Diagnostic plots in Cox's regression model. Biometrics 1991; 47: 841-850.

Cnaan A, L Ryan. Survival analysis in natural history studies of disease. Statist Medicin 1989; 8: 1255-1268.

Cox DR, EJ Snell. A general definition of residuals. J Roy Statist Soc 1968; B30: 248-275.

Cox DR. Regression models and life tables (with discussion). J R Statist Soc 1972; B34: 187-220.

Cox DR. Partial likelihood. Biometrika 1975; 62: 269-276.

Cox DR. The analysis of binary data. 1970. Methuen, London.

Cox DR, DV Hinkley. Theoretical statistics. Chapman and Hall, London 1974.

Cox DR, D Oakes. Analysis of survival data. Chapman and Hall, London 1984.

Damhuis RAM, JHM Blom. The influence of age on treatment choice and survival in 735 patients with renal carcinoma. Br J Urol 1995; 75: 143-147.

Farewell VT, RL Prentice. A study of distributional shape in life testing. Technometrics 1977; 19: 69-76.

Farewell VT, B Math, M Math. The combined effect of breast cancer risk factors. Cancer 1977; 40: 931-936.

Feigl P, M Zelen. Estimation of exponential survival probabilities with concomitant information. Biometrics 1965; 21: 826-838.

Feinleib M, B MacMahon. Variation in the duration of survival of patients with chronic leukemias. Blood 1960; 332-349.

Finkelstein DM. A proportional hazards model for interval-censored failure time data. Biometrics 1986; 42: 845-854.

Fleming TR, JR O'Fallon, PC O'Brien, DP Harrington. Modified Kolmogorov-Smirnov test procedures with application to arbitrarily right censored data. Biometrics 1980; 36: 607-626.

Fossa SD, A Kramar, JP Droz. Prognostic factors and survival in patients with metastatic renal cell carcinoma treated with chemotherapy or interferon-alpha. Eur J Cancer 1994; 30A: 1310-1314.

Gill RD. Censoring and stochastic integrals. Mathematical Centre Tracts 124. Mathematisch Centrum, Amsterdam, 1980.

Greenwood M. The natural duration of cancer. Reports on public health and medical subjects 1929; vol 33: 1-26.

Gross AJ, VA Clark. Survival distributions: reliability applications in the biomedical sciences. Wiley, New York, 1975.

Hall WJ, JA Wellner. Confidence bands for a survival curve from censored data. Biometrika 1980; 67: 133-143.

Harrington DP, TR Fleming. A class of rank test procedures for censored survival data. Biometrika 1980; 69: 553-566.

Johnson ME, HD Tolley, MC Bryson, AS Goldman. Covariate analysis of survival data: a small-sample study of Cox's model. Biometrics 1982; 38: 685-698.

Kalbfleisch JD, RL Prentice. The statistical analysis of failure time data. Wiley, New York, 1980.

Kaplan EL, P Meier. Nonparametric estimation from incomplete observations. J Am Statist Assoc 1958; 53: 457-481.

Kay R. Proportional hazard regression models and the analysis of censored survival data. Appl Statist 1977; 26: 227-237.

Kay R. Treatment effects in competing risks analysis of prostate cancer data. Biometrics 1986; 42: 203-211.

Kellerer AM, D Chmelevsky. Small-sample properties of censored data rank tests. Biometrics 1983; 39: 675-682.

Koziol JA. Goodness-of-fit tests for randomly censored data. Biometrika 1980; 67: 693-696.

Kurtzke JF. On estimating survival: A tale of two censors. J Clin Epidemiol 1989; 42: 169-175.

Lagakos SW, DA Schoenfeld. Properties of proportional hazards score tests under misspecified regression models. Biometrics 1984; 40: 1037-1048.

Lee ET. Statistical methods for survival data analysis. Lifetime Learning Publications, Belmont California, 1980

Lehmann EL. The power of rank tests. Ann Math Statist 1953; 24: 23-43.

Lin DY, TR Fleming. Confidence bands for survival curves under the proportional hazards model. Biometrika 1991

Link CL. Confidence intervals for the survival function using Cox's proportional hazards model with covariates. Biometrics 1984; 40: 601-610.

Lopez Hanninen E, H Kirchner, J Atzpodien. Interleukin-2 based home therapy of metastatic renal cell carcinoma; risks and benefits in 215 consecutive single institution patients. J Urol 1996; 155: 19-25.

Mantel N. Evaluation of survival data and two new rank order statistics arising in its consideration. Cancer Chemother Rep 1966; 50: 163-170.

Mantel N, W Haenszel. Statistical aspects of the analysis of data from retrospective studies of disease. J Nat Cancer Inst 1959; 22: 719-748.

Mantel N, M Myers. Problems of convergence of maximum likelihood iterative procedures in multiparameter situations. J Am Statist Asssoc 1971; 335: 484-491.

Metcalf W. Analysis of cancer survival as an exponential phenomenon. Surgery Gynaec Obstet. 1974; 138: 731-740.

Miller RG. Survival analysis. Wiley, New York, 1981.

Nelson W. Hazard plotting for incomplete failure data. J Qual Technol 1969; 1: 27-52.

O'Quigley J, F Pessione. Score tests for homogeneity of regression effect in the proportional hazards model. Biometrics 1989; 45: 135-144.

Peto R. Rank tests of maximal power against Lehmann-type alternatives. Biometrika 1972; 2: 472-475.

Peto R, PN Lee, WS Paige. Statistical aspects of the bioassay of continuous carcinogens. Br J Cancer 1972; 26: 258-261.

Pocock SJ, SM Gore, GR Kerr. Long term survival analysis: the curability of breast cacer. Statist Medic 1982; 1: 93-104.

Prentice RL. Exponential survivals with censoring and explanatory variables. Biometrika 1973; 60: 279-288.

Prentice RL, LA Gloeckler. Regression analysis of grouped survival data with application to breast cancer data. Biometrics 1978; 34: 57-67.

Prentice RL, P Marek. A qualitative discrepancy between censored data rank tests. Biometrics 1979; 35: 861-867.

Ramlau-Hansen H. Smoothing counting process intensities by means of kernel functions. Ann Stat 1983; 11: 453-466.

Roosen JU, U Engel, RH Jensen, E Kvist, G Schou. Renal cell carcinoma: prognostic factors. Br J Urol 1994; 74: 160-164.

Schoenfeld D. Chi-squared goodness-of-fit tests for the proproportional hazards model. Biometrika 1980; 67: 145-153.

Schwab RJ. Disturbances of sleep in the intensive care unit. Crit Care Clin 1994; 10: 681-694.Thomas

Therneau TM, PM Grambsch, TR Fleming. Martingale-based residuals for survival models. Biometrika 1990; 77: 147-160.

Thomas DR, GL Grunkemeier. Confidence interval estimation of survival probabilities for censored data. J Am Statist Assoc 1975; 70: 865-871.

Thomas DG, N Breslow, JJ Gart. Trend and homogeneity analyses of proportion and life table data. Comput Biomed Res 1977; 10: 373-381.

Tsiatis AA. A large sample study of Cox's regression model. Ann Statist 1981; 9: 93-108.

Whitmore AS, JB Keller. Survival estimation using splines. Biometrics 1986; 42: 495-506.

Zelen M. Applications of exponential models to problems in cancer research. J Roy Statist Soc 1966; A3: 368-398.

Zippin C, P Armitage. Use of concomitant variables and incomplete survival information with estimation of an exponential survival parameter. Biometrics 1966; 22: 665-672.

CHAPTER 3

PROGNOSTIC INDEXES FOR RENAL CELL CARCINOMA

This chapter was published in European Journal of Cancer 1980; 16: 833-840
by W.C.J. Hop and B.H.P. van der Werf-Messing.

Introduction

As in many other malignant diseases, patients with adenocarcinoma of the kidney show a great variability in survival as the outcome of treatment. The identification of factors associated with the prognosis of these patients, and from which the variability in outcome can be substantially explained, may be of great importance. In this paper, results of a study to investigate the relative prognostic importance of a number of factors concerning nonmetastasized renal cell carcinoma by a multivariate statistical method are presented. Two analyses are performed. One is concerned with preoperatively assessable factors. In the second, tumour characteristics which become evident on histopathological examination of the nephrectomy specimen are also included. This distinction is useful, as the preoperative identification of groups of patients with different prognosis may facilitate treatment decisions.
Postoperatively, emphasis is on explaining variability in treatment results by consideration of possible mechanisms of the disease.
The preoperative factors studied are (a) sex, (b) age, (c) erythrocyte sedimentation rate (ESR) and (d) radiodiagnostic extent of the tumour. Postoperative tumour characteristics analyzed are (e) histopathological extent of the tumour, (f) invasion of the renal vein, (g) degree of differentiation and (h) cell type. In both analyses, prognostic indexes from which the expected survival of individual patients can be derived are constructed.

Material and methods

During the years 1965-1977, 174 patients with nonmetastasized renal cell carcinoma who were considered suitable for simple nephrectomy entered a randomized clinical trial to investigate the value of preoperative irradiation[1]. Of these patients, 89 received preoperative irradiation (TD 3000-4000 rad/3-4 weeks) and 85 were directly nephrectomized.
The i.v. pyelographs and arteriographs performed for the preoperative assessment of the extent of the tumour, the T-category[2], were reviewed by one diagnostic radiologist (R.C. Ledeboer). Renal vein invasion was assessed by histological examination of the nephrectomy specimen. The histological extent of the tumour is expressed by the P-category (P1: tumour surrounded by renal parenchyma; P2: tumour extending to the capsule and/or invading the renal pelvis and/or calyces; P3: perinephric or hilar extension; P4: extension into neighbouring organs and/or fixed to the abdominal wall). Review of all histological specimens was done by one pathologist (R.O. van der Heul). Because of simple nephrectomy local node involvement was not assessed as a routine.
At the time of this analysis, 87 patients had died and 87 were still continuing follow-up. All patients had regular follow-up examinations.

Statistical methods

Survival functions of groups of patients are estimated by survival curves according to Kaplan and Meier[3]. For the calculation of P-values in the comparison of survival curves, the log-rank test[4] is used. Because of the limited number of patients and the existing relationships among several factors, the prognostic importance of the factors can be simultaneously investigated by these methods to only a limited extent. A more promising way of making progress here is by adopting a multivariate statistical model. Moreover, such a model may be useful in deriving a rule for prognostic predictions concerning future individual patients. The model which is used here is Cox's proportional hazards model[5]. With this model, a scoring function which relates the expected survival times of individual patients to the values of the prognostic variables can be obtained. The value of the scoring function indicates how strongly the force of mortality (the instantaneous death rate) is related to these factors.
If the number of patients is denoted by n and, if p is the number of prognostic variables, the scoring function for the i-th patient can be written:

$$S_i = a_{i1}\beta_1 + a_{i2}+\beta_2 + \ldots + a_{ip}\beta_p. \qquad (1)$$

Here a_{ij} is the value of the j-th prognostic variable ($j=1,\ldots,p$)for the i-th patient ($i=1,\ldots,n$).The parameter β_j denotes the difference in score of patients who differ by one unit in the j-th variable, the other variables being at the same level. Factors with continuous levels such as ESR and discrete factors with two categories such as sex are represented by one variable and one related parameter β. A discrete factor with more than two categories generally needs a number of variables and parameters equal to its number of categories minus one. For instance, the four-categorical factor T-category needed three variables to represent all its levels.
The score S_i is related to the survival outcome by:

$$\lambda_i(t) = \exp(S_i)\,\lambda_0(t),$$

where $\lambda_i(t)$ denotes the force of mortality of the i-th patient at time t (roughly speaking, $\lambda_i(t)$ gives the probability of dying at month t if the patient is known to be alive at month t-1). $\lambda_0(t)$ denotes an arbitrary reference force of mortality function which is unknown. A high score indicates a high death rate and thus a (relatively) poor prognosis. Another interpretation of this scoring system can be obtained by considering its implication for the survival functions. It can be shown that the survival function $F_i(t)$ to which the i-th patient is subjected can be written as:

$$F_i(t) = \{F_0(t)\}^{\exp(Si)}.$$

Here $F_i(t)$ denotes a reference survival function, that is, the survival function for patients with a score $S=0$.
The unknown parameters $\beta_1, \beta_2, \ldots, \beta_p$ and the function $F_0(t)$ can be estimated from the clinical data by maximum likelihood procedures[6]. This fitting of the model has to be done by a process of successive approximation. Significance tests on subsets of less important prognostic factors can be obtained by comparisons of fits of models through likelihood ratios.

Results

Extensive analysis of the data showed that the preoperative irradiation had little or no effect on histological variables or survival (5-year survivals for the preoperatively treated and the surgically treated only group were, respectively, 45 and 54%), This treatment factor is therefore omitted from further consideration. Preoperative factors were first analyzed for their prognostic value.

Table 1. Survival according to preoperative factors.

Factor	Category	No. of patients	Survival (%) 3yr	5yr	Log-rank test
Sex	female	75	66	57	
	male	99	59	43	$P=0.05$
Age (yr)	<=55	60	65	60	
	56-65	60	59	38	
	>=66	54	61	50	$P=0.2$
Sedimentation rate (mm/hr)	<=12	50	89	80	
	13-29	39	76	57	
	>=30	83	40	27	$P<0.001$
T-category	T1	14	75	48	
	T2	47	58	50	
	T3	66	70	57	
	T4	39	52	36	$P=0.1$
combined	T1,2,3	127	66	52	
	T4	39	54	36	$P=0.02$

Missing data: sedimentation rate: 2, T-category: 8.

In Table 1, 3-year and 5-year survival percentages are given for the preoperative assessable factors. Considered on its own, sedimentation rate seems to be strongly associated with subsequent survival. Females have a better prognosis than males. Category T4 does more poorly than the lower T-categories. To analyse the effects of the factors sedimentation rate, sex and T-category simultaneously, model (1) is written as:

$$S_i = a_{i1}\beta_1 + a_{i2}+\beta_2 + a_{i3}+\beta_3 + a_{i4}+\beta_4 + a_{i5}+\beta_5.$$

Here a_{ij} allows for the factor sedimentation rate and a_{i2} for sex. The three variables a_{i3}, a_{i4}, and a_{i5} represent the four categories of the T-classification. Further details on the coding of these variables and the resulting fit to the data are given in Table 2. It

Table 2. Model including sedimentation rate, sex and T-category.

Factor	Variable	ß	Estimated ß	Likelihood ratio test
Sedimentation rate (mm/hr)	a_{i1} = ESR	$ß_1$	0.020	$P<0.001$
Sex	a_{i2} = 1 if male = 0 if female	$ß_2$	0.51	$P=0.05$
T-category	a_{i3} = 1 if T1 = 0 if otherwise	$ß_3$	-0.26	$P=0.01$
	a_{i4} = 1 if T2 = 0 if otherwise	$ß_4$	-0.39	
	a_{i5} = 1 if T3 = 0 if otherwise	$ß_5$	-0.95	

appears that each of these three factors contributes significantly to the score. On further examination of the estimated parameters related to the T-category, however, it appears that category T3 has a more favourable prognosis than T2. In turn, T2 is estimated to be somewhat more favourable than T1. As this ranking did not seem to be logical and the differences between the categories T1, T2 and T3 themselves were far from statistical significance, the model was fitted again to the data with T4 contrasted to categories T1, T2 and T3 combined. After assuring that the factor age did not improve the fit of the model, the ultimate preoperative score could be written as:

$$S = 0.02 \times \text{ESR} + \begin{cases} 0.6 \text{ (male)} \\ 0 \text{ (female)} \end{cases} + \begin{cases} 0.7 \text{ (T4)} \\ 0 \text{ (T1, T2 or T3)} \end{cases}$$

Now, for each patient, a score according to this prognostic index can be calculated and patients with a similar score can be grouped. After grouping patients with a score between 0 and 1, between 1 and 2, etc., it is shown in Figure 1 that the observed survival curves of these groups of patients are in reasonably good agreement with the expected survival curves arising from the model. Groups with a relatively good, intermediate and poor prognosis can be clearly distinguished.

Concerning postoperative factors, it appeared that category P4 (inoperable patients) had a very poor prognosis. All seven patients died within 2 yr, most of them within 1 yr. Moreover, as no information was available concerning the histological variables of the tumour in the majority of these patients, they were omitted from the following analysis. In Table 3, survival percentages are given for the remaining operable patients. With increasing P-category, survival becomes worse. Renal vein involvement

Table 3. Survival according to histopathological factors

Factor	Category	No. of patients	Survival (%) 3 yr	5 yr	Log-rank test
P-category	P1	61	81	65	
	P2	38	64	60	
	P3	67	50	33	$P=0.001$
Renal vein invasion	Yes	62	47	28	
	No	105	76	65	$P<0.001$
Degree of differentiation	Low	60	51	39	
	Medium	81	73	57	
	High	21	90	68	$P=0.01$
Cell type	Clear	123	68	52	
	Granular	17	67	52	
	Mixed	22	54	54	$P=0.8$

Missing data: P-category: 1, degree of differentiation and cell type: 5.

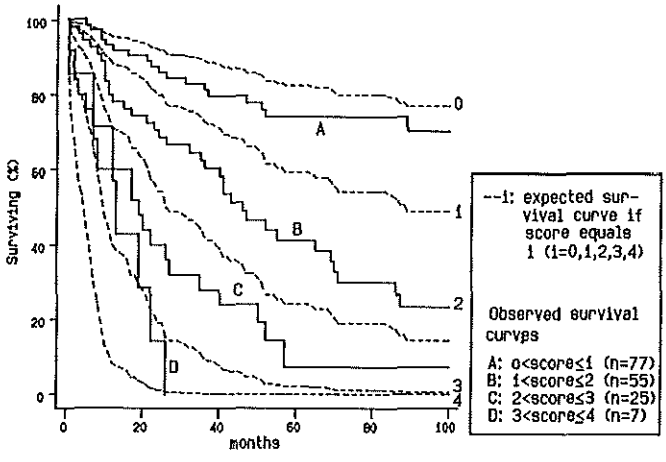


Figure 1. Observed and expected survival curves according to preoperative score [=0.02 x ESR (mm/hr) + 0.6 (if male) + 0.7 (if T-category T4)].

leads to a considerably poorer prognosis than when no venous invasion is seen. Also with decreasing degree of differentiation of the tumour, survival becomes worse, while the different cell types show similar survival rates. To investigate the prognostic importance of the P-category, renal vein invasion and degree of differentiation simultaneously, the following scoring function, which also takes account of sex and sedimentation rate, was fitted to the data:

$$S_i = a_{i1}\beta_1 + a_{i2}+\beta_2 + \ldots + a_{i7}\beta_7.$$

The various variables and results of fitting this model are given in Table 4. It appears that, once the factors sex, sedimentation rate and renal vein involvement are known, the factors P-category and degree of differentiation do not contribute significantly to the score. Extending the model with the factors age and cell type also did not improve the fit of the model. The ultimate postoperative score, based on all operable patients except two because of a missing sedimentation rate, can now be written as:

$$S = 0.02 \times \mathrm{ESR} + \begin{cases} 0.8 \text{ (male)} \\ 0 \text{ (female)} \end{cases} + \begin{cases} 0.9 \text{ (renal vein invasion)} \\ 0 \text{ (no renal vein invasion)} \end{cases}$$

After grouping patients with a similar score, Figure 2 is obtained. Good agreement between observed and expected survival curves is evident.

Table 4. Model including sedimentation rate, sex, P-category, renal vein invasion and degree of differentiation.

Factor	Variable	ß	Estimated ß	Likelihood ratio test
Sedimentation rate	a_{i1}=ESR	β_1	0.018	$P<0.001$
Sex	a_{i2}=1 if male =0 if female	β_2	0.87	$P=0.001$
P-category	a_{i3}=1 if P1 =0 if otherwise	β_3	-0.48	$P=0.15$
	a_{i4}=1 if P2 =0 if otherwise	β_4	-0.53	
Renal vein invasion	a_{i5}=1 if yes =0 if no	β_5	-0.95	$P=0.002$
Degree of differentiation	a_{i6}=1 if low =0 if otherwise	β_6	0.25	$P=0.3$
	a_{i7}=1 if medium =0 if otherwise	β_7	-0.14	

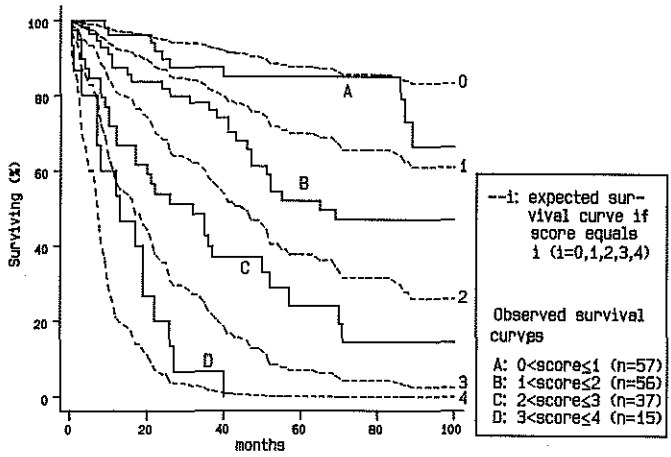


Figure 2. Observed and expected survival curves according to postoperative score [=0.02 x ESR (mm/hr) + 0.8 (if male)+ 0.9 (if renal vein involvement)].

Discussion

The relatively poor prognosis of patients with invasion of the renal vein is in agreement with general clinical experience[7]. The risk of renal vein involvement increases greatly with increasing P-category; see Table 5. It is conceivable that the prognostic importance of P-categories is mainly derived from renal vein involvement. A similar phenomenon could be observed in the data of McDonald and Priestley[8], who studied survival in relation to renal vein involvement and weight of the diseased kidney.

Table 5. Numbers of patients according to renal vein invasion and P-category

		P-category		
		P1	P2	P3
Renal vein invasion	Yes	8 (13%)	15 (39%)	39 (58%)
	No	53	23	28

Chi-square test: $P<0.001$.

It has been known for a long time that renal cell carcinoma patients frequently show an elevated sedimentation rate[9]. The importance of the sedimentation rate as a prognostic factor was already stated by Böttiger[10], Ochsner et al.[11] and recently by Juusela[12]. In this study, a gradual worsening of prognosis was observed with increasing sedimentation rate. The sedimentation rate showed a strong correlation with the degree of differentiation. With a decreasing degree of differentiation, the sedimentation rate increases. Cumulative frequency distributions of the sedimentation rate are

given in Figure 3 (differences significant at P=0.01 by the Krukal-Wallis test). This correlation could explain why the degree of differentiation, while of prognostic value when considered alone (Table 3), is of less value in the postoperative score.

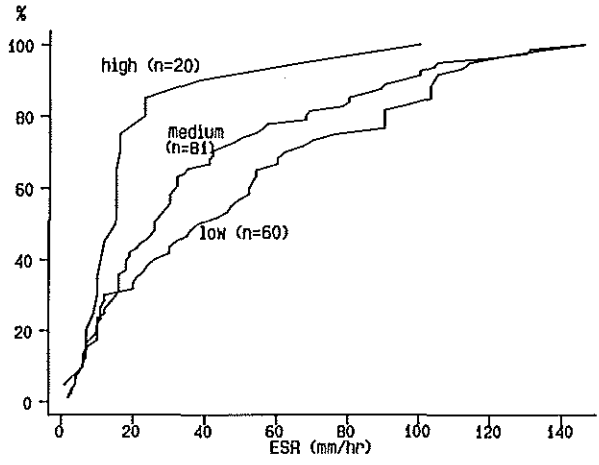


Figure 3. Cumulative frequency distributions of erythrocyte sedimentation rate by degree of differentiation.

That females have a better prognosis than males was also noted by Meyers et al[13] and Juusela[12]. The greater incidence of renal cell carcinoma among males[14] and the finding that, once metastases have become clinically evident, women respond less to hormonal treatment than men[15,16] are also suggestive of the involvement of hormonal factors.
The prognostic value of the T-classification was not impressive in this material. Only category T4 was an ominous sign. This observation could readily be explained by comparing T with P-categories. While T4 patients were mainly P3 or P4, there was a minor agreement between T and P-categories for the cases with a lower T-category.
The prognostic indexes derived were broadly similar when a correction was made for the occurrence of intercurrent death. By considering patients (18) withdrawn from study at the moment of death from causes supposedly unrelated to renal cell carcinoma, the estimated parameters in the multivariate models did not appreciably differ from those given.
In comparing observed and expected survival curves (Figures 1 and 2), good agreement was apparent. The validity of a prognostic index, however, should preferably be tested in another group of comparable patients. Dr. H. Juusela (Helsinki University Central Hospital) kindly evaluated the prognostic indexes by applying them to a comparable group of patients described in his thesis[12]. For patients without evidence of disseminated disease and with a P-category less than 4, intercurrent death corrected survival curves according to postoperative score were obtained as given in Figure 4 (one patient with a score of 4.4 dying 1 month after operation is not included in that figure). Despite the fact that the groups with the highest scores are less separated, the overall worsening of prognosis with increasing score is satisfactorily confirmed. For the preoperative index a similar validation was obtained.
About uses of the indexes: the preoperative prognostic indicator may possibly be

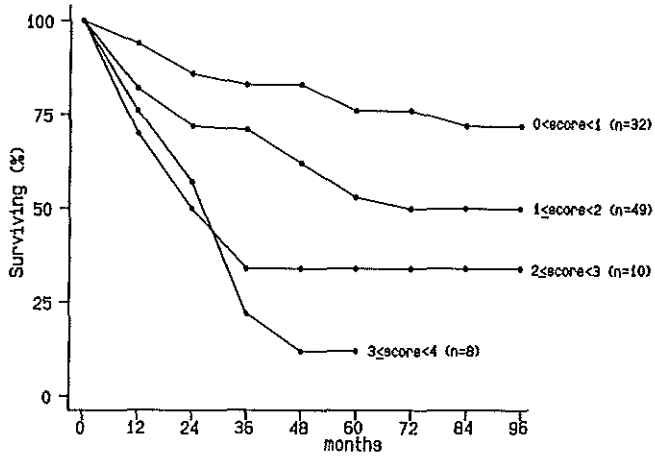


Figure 4. Survival after extrafascial nephrectomy. Patients are grouped according to postoperative score (courtesy of Dr. H. Juusela).

useful as a complementary factor to facilitate making the decision of whether to operate or not in particular patients. Until now, no mode of chemotherapy has been found which can deel effectively with clinically established metastases. In studies of more aggressive hormonal of chemotherapy treatments, the indexes can be useful either as selection criteria or as single stratification factors.

References

1. Werf-Messing B van der, RO van der Heul, RC Ledeboer. Renal cell carcinoma trial. Cancer Clin Trials 1978; 1: 13.
2. Union Internationale Contre le Cancer (UICC). TNM Classification of Malignant Tumours (2nd ed.), Geneva, 1974.
3. Kaplan EL, P Meier. Nonparametric estimation from incomplete observation. J Amer Statist Ass 1958; 53, 457.
4. Peto R, MC Pike, P Armitage, NE Breslow, DR Cox, SV Howard, N Mantel, K McPherson, J Peto, PG Smith. Design and analysis of randomized clinical trials requiring prolonged observation of each patient. Brit J Cancer 1977; 35: 1.
5. Cox DR. Regression models and life tables (with discussion). J Roy Stat Soc 1972; B34: 187.
6. Breslow N. Analysis of survival data under the proportional hazards model. Int Stat Rev 1975; 43: 44.
7. Riches E. Tumours of the Kidney and Ureter. Livingstone, Edinburgh, 1964.
8. McDonald JR, JT Priestley. Malignant tumours of the kidney: surgical and prognostic significance of tumour thrombosis of the renal vein. Surg Gynec Obstet 1943; 77: 295.
9. Böttinger LE. Studies in renal carcinoma: clinical and pathologic anatomical aspects. Acta Med Scand 1960; 167: 443.

10. Böttinger LE. Prognosis in renal carcinoma. Cancer(Philad.) 1970; 26: 780.
11. Ochsner MG, W Brannan, HS Pond, EH Goodier. Renal cell carcinoma: review of 26 yr of experience at the Ochsner clinic, J Urol 1973; 110: 643.
12. Juusela H. Renal adenocarcinoma, clinical features and factors correlating to survival in 173 patients. Ann Chir Gynaec 1978; 67 (suppl 193).
13. Myers GH, LG Fehrenbaker, PP Kelalis. Prognostic significance of renal vein invasion by hyper nephroma. J Urol 1968: 100; 420.
14. Bennington JL, RM Kradjan. Renal Carcinoma. Saunders, Philadelphia, 1967.
15. Bloom HJG. Hormone treatment of renal tumours: experimental and clinical observations. In: Tumours of the Kidney and Ureter. (Eds. E. Riches) Vol. 5. Livingstone, Edinburgh,- 1964.
16. Werf-Messing B van der, HA van Gilse. Hormonal treatment of metastases of renal carcinoma. Brit J Cancer 1971; 25: 423.

CHAPTER 4

BLOOD TRANSFUSIONS AND PROGNOSIS IN COLORECTAL CANCER

This chapter was published in The New England Journal of Medicine 1993; 328: 1372-1376
by O.R.C. Busch, W.C.J. Hop, M.A.W. Hoynck van Papendrecht, R.L. Marquet and J. Jeekel.

Introduction

Perioperative blood transfusions may have a deleterious effect on the survival of patients with a variety of solid tumors[1,2], possibly because of an immunosuppressive effect[3,4]. This possibility is supported by studies in animals in which tumor growth was enhanced after allogeneic transfusion[5,6], although conflicting results have also been reported[7,8]. A poor prognosis after blood transfusions has been noted especially in patients with colorectal cancer.
In the studies of the effect of blood transfusions in patients with cancer, the patients were given the transfusions either because of their disease or because transfusion was necessary during surgical treatment. Therefore, the question remains whether the relation between blood transfusion and poor prognosis is causal or coincidental[9]. The need for transfusion could be an indicator of other prognostic factors that are either unknown or difficult to quantify, such as the extent of the tumor and the dissection, the skill of the surgeon, and the nutritional state of the patient.
A randomized trial is the only way in which possible bias in the selection of patients can be avoided. Randomization between transfusion and no transfusion is impossible, however, because giving patients transfusions when there is no medical indication and withholding transfusions that are indicated are ethically unacceptable. Since autologous blood is the safest blood to use in transfusion, comparing the effects of allogeneic and autologous blood transfusions would be a logical option.
We therefore conducted a randomized multicenter trial in patients with colorectal cancer to determine whether autologous blood transfusion would reduce the rate of recurrence of cancer and improve survival as compared with allogeneic transfusion. We previously reported that a notable reduction in the number of allogeneic transfusions can be achieved with a program of autologous blood transfusion[10].

Methods

The study was conducted in 14 hospitals in the Netherlands and 1 hospital in England and was approved by the ethics committees of all the participating hospitals. After written informed consent had been obtained, eligible patients were randomly assigned to either the allogeneic-transfusion group or the autologous-transfusion group, with stratification according to the participating hospital. Patients were enrolled from August 1986 to November 1991, when the planned enrollment was reached.

Eligibility of patients:
Patients scheduled for a potentially curative resection of cancer of the colon or rectum were eligible for enrollment if they fulfilled the criteria of the American Association of Blood Banks for autologous blood donation in anticipation of surgery[11]. These criteria required the absence of severe cardiovascular or respiratory disease, no history of epilepsy after infancy, and a hemoglobin concentration above 11.3g per deciliter (7 mmol per liter). In addition, patients had to have no evidence of metastatic disease, on the basis of chest radiography and ultrasonography of the liver, no other cancer except basal-cell carcinoma of the skin or in situ carcinoma of the cervix; no evidence of ulcerative colitis, familial polyposis, or a fixed rectal carcinoma requiring preoperative radiation therapy; and no history of blood transfusion during the three months before randomization. No adjuvant therapy was allowed except irradiation. If metastatic or recurrent disease developed in a patient during follow-up, all available therapies were allowed.

Procedures for donation and transfusion:
Patients randomly assigned to the autologous-transfusion group were required to donate blood twice. The minimal interval between the two donations was 72 hours, and the second donation had to occur not later than five days begore surgery. At each donation, 450 ml of blood was obtained by standard procedures.
The patients were treated with oral iron supplementation immediately after randomization.
The collected blood was separated into packed red cells and fresh-frozen plasma, except at one hospital, where autologous blood was given in transfusion as whole blood. The packed red cells and whole blood were stored at 4°C. Allogeneic blood was always given in transfusion as packed red cells. Standard rules for transfusion were used for both groups. Packed red cells could be given only if the loss of blood exceeded 500 ml or if the hemoglobin concentration dropped below 10.5g per deciliter (6.5 mmol per liter). If this hemoglobin concentration was not achieved after two autologous transfusions, additional allogeneic transfusions were made. In both groups fresh-frozen plasma was given when indicated.

Surgery and histopathological assessment:
Standard surgical procedures were used. The operative specimens were classified according to Dukes' classification as modified by Turnbull[12]. A tumor confined to the bowel wall was classified as Dukes' stage A, a tumor extending through the serosa into the pericolic fat as Dukes' stage B, the presence of regional lymph nodes containing metastases as Dukes' stage C, and the presence of distant metastases or unresectable tumor as Dukes' stage D. All patients who had residual tumor evident only on microscopical examination received postoperative radiotherapy. These patients and those who had en bloc resection of adjacent organs were not considered as having Dukes' stage D disease, but rather as having Dukes' stage B or C disease.

Follow-up and criteria for recurrent cancer:
The patients were evaluated every three months during the first two years after surgery and every six months thereafter. Each evaluation consisted of a history, a physical examination, and blood tests (to measure the concentrations of hemoglobin and serum carcinoembryonic antigen). Ultrasonography of the liver was performed

every six months for three years and each year thereafter. Chest films and colonoscopy were done yearly.
Characteristic abnormalities detected on physical examination or on chest radiography, liver ultrasonography, or abdominal computed tomography were accepted as evidence of metastatic or recurrent disease. If possible, the presence of metastatic or recurrent disease was confirmed by histologic or cytologic examination. Increased serum concentrations of carcinoembryonic antigen without evidence of recurrence at suspected anatomical sites were not considered to indicate metastatic or recurrent disease.

Statistical analysis:
Categorial data were compared by the chi-square test, and continuous data by the Mann-Whitney test. The major end points were disease-free survival and colorectal cancer-specific survival (defined to include all patients who did not die of colorectal cancer, without regard to other causes of death), both as determined from the time of surgery and calculated according to the Kaplan-Meier method. The log-rank test was used to compare these end points. Multivariate analysis was performed by proportional-hazards analysis[13] to obtain a higher level of precision in the comparison of the transfusion groups. Two-sided P values of 0.05 or less were considered statistically significant. P values calculated with adjustment for Dukes' stage are indicated as adjusted P values in the text.
Patients in whom a second primary tumor developed outside the colon were considered withdrawn from the study with regard to the calculation of disease-free survival. Metachronous tumors in the colon, however, were defined as representing recurrent disease. Postoperative deaths, defined as deaths occurring within 30 days after surgery, and deaths occurring more than 30 days after surgery due to postoperative complications were counted as deaths due to cancer.
Because the analyses of overall survival and colorectal cancer-specific survival gave similar results, only the results for colorectal cancer-specific survival are reported.
Except for the patients who did not have colorectal cancer at the time of surgery, all randomized patients were primarily evaluated according to the intention-to-treat principle. In addition, exploratory analyses were performed according to the number and type of transfusions received.

Results

Characteristics of the Patients:
A total of 510 patients were enrolled in the study. Thirty-five patients (7 percent) were excluded because they did not have colorectal cancer at the time of surgery. The characteristics of the remaining 475 patients are shown in Table 1. None of the characteristics differed significantly between the two groups. Twenty-six of the 239 patients in the autologous-transfusion group (11 percent) did not donate blood; the majority of these were refused by the blood bank because they did not fulfil the criteria of the American Association of Blood Banks. These patients were included in all the analyses according to the intention-to-treat principle, as were all the patients with metastatic disease or unresectable tumors. The median follow-up period was 2.5 years (range, 1 to 59 months), and no patient was lost to follow-up.
The perioperative hematologic values and use of transfusions are shown in Table 2.

Table 1. Characteristics of the patients with colorectal cancer in the two transfusion groups.

Characteristic	Allogeneic Transfusion (N = 236)	Autologous Transfusion (N = 239)
Median age - yr (range)	68 (33-89)	66 (31-88)
	number (percent)	number (percent)
Sex		
Male	132 (56)	141 (59)
Female	104 (44)	98 (41)
Tumor location		
Ascending colon	24 (10)	16 (7)
Flexures and transverse colon	12 (5)	16 (7)
Descending colon and sigmoid	62 (26)	65 (27)
Rectosigmoid and rectum	132 (56)	135 (56)
Multiple primary tumors	6 (3)	7 (3)
Dukes' classification*		
A	53 (23)	55 (23)
B	85 (36)	80 (34)
C	78 (33)	72 (31)
D	18 (8)	29 (12)
Histologic differentation		
Good	35 (15)	33 (14)
Moderate	169 (72)	172 (72)
Poor	29 (12)	29 (12)
Unknown	3 (1)	5 (2)
Adjacent-organ fixation°	15 (7)	17 (8)
Adjuvant irradiation°	15 (7)	16 (8)

* Does not include patients who did not undergo surgery (two in the allogeneic group and three in the autologous group)

° Excludes patients with Dukes' stage D tumors.

The number of patients in the autologous-transfusion group who received allogeneic blood was half the number in the allogeneic-transfusion group (28 percent vs. 56 percent, respectively; $P<0.001$).

Morbidity and Mortality:

Five eligible patients (two in the allogeneic-transfusion group and three in the autologous-transfusion group) did not have surgery. Four had more advanced disease than expected, and one patient died before surgery. Eight patients (three in the allogeneic-transfusion group and five in the autologous-transfusion group) died of postoperative complications. Postoperative infectious complications occurred in 26 percent of the patients (25 percent in the allogeneic-transfusion group and 27 percent

in the autologous-transfusion group). There were no statistically significant differences between the two groups with respect to postoperative mortality and infectious complications.

Disease-free Survival:
Of the 423 patients who underwent curative surgery (216 in the allogeneic-transfusion group and 207 in the autologous-transfusion group), 105 (54 in the allogeneic-transfusion group and 51 in the autologous-transfusion group) had recurrent disease (including three metachronous tumors). The disease-free survival of these 423 patients is shown in Figure 1. The plots for the two groups were almost identical (P=0.93).

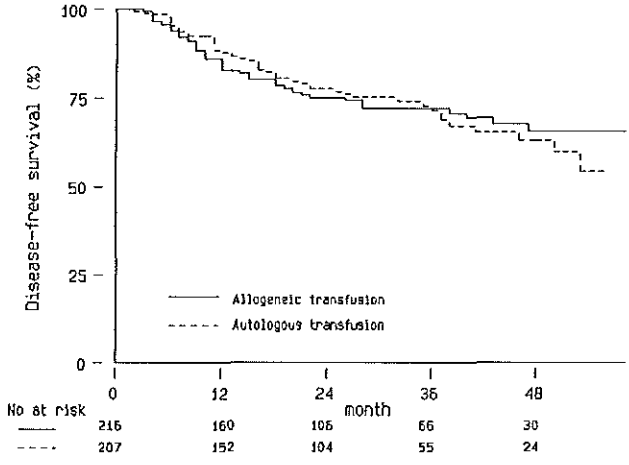


Figure 1. Disease-free survival of all 423 patients with colorectal cancer who underwent curative surgery. The disease-free survival at four years was 66 percent in the allogeneic transfusion group and 63 percent in the autologous transfusion group (P=0.93). The 95 percent confidence interval for the difference between these percentages (autologous minus allogeneic) ranged from -16 percent to + 10 percent.

Also, there were no differences with regard to diseasefree survival in each of the Dukes' stages. After adjustment for various factors, the multivariate analysis also revealed no difference between the groups in disease-free survival (Table 3). Six patients (two in the allogeneic-transfusion group) had second primary tumors outside the colon during follow-up.

To explore the relation between blood transfusions and disease-free survival, we grouped the patients who underwent curative surgery according to the number and type of transfusions they received. The disease-free survival in the 143 patients who received no transfusions was significantly better (adjusted P=0.001) than that in the 280 who did receive transfusions; at four years, it was 73 percent and 59 percent, respectively. Among the 280 patients who had transfusions, 136 received only alloge-

Table 2. Hemoglobin concentrations, blood loss, and transfusions in patients with colorectal cancer, according to transfusion group.

Variable	Allogeneic Transfusion	Autologous Transfusion	P value
	median (range)		
Hemoglobin concentration (g/dl)			
At baseline	14.5 (15.0-18.0)	14.4 (10.7-18.5)	NS*
Immediately before surgery	14.1 (9.5-18.0)	12.5 (8.4-16.4)	<0.001
At discharge	12.5 (9.2-17.2)	12.2 (9.2-16.2)	NS
Blood loss (ml)	775 (100-11,500)	750 (100-6500)	NS
Transfusions			
None	103 (44)	61 (26)	<0.001
Autologous only	- (-)	112 (47)	-
Allogeneic	113 (56)	66 (28)	<0.001

* NS denotes not significant.
To convert values for hemoglobin to millimoles per liter, multiply by 0.62.

neic transfusions, 102 only autologous transfusions, and 42 transfusions of both types. The disease-free survival at four years in these three groups was 56 percent, 62 percent, and 66 percent, respectively (adjusted P=0.50). No significant difference was found between the 49 patients who received no transfusions in the autologous-transfusion group and the 94 such patients in the allogeneic-transfusion group; the disease-free survival in these patients at four years was 69 percent and 75 percent, respectively.

Because the number of autologous transfusions was limited to two, further comparisons were made between the 75 patients in the allogeneic-transfusion group who received one or two allogeneic transfusions and the 102 patients in the autologous-transfusion group who received one or two autologous transfusions but no allogeneic transfusions. The disease-free survival was significantly worse in the patients in both groups who received transfusions than in the 143 patients who did not, whereas the disease-free survival of the patients who received transfusions in both groups did not differ significantly from each other (Table 4). Analysis of the results with respect to the use of fresh-frozen plasma revealed no relation between its use and disease-free survival. This was true whether the patients received blood transfusions or not.

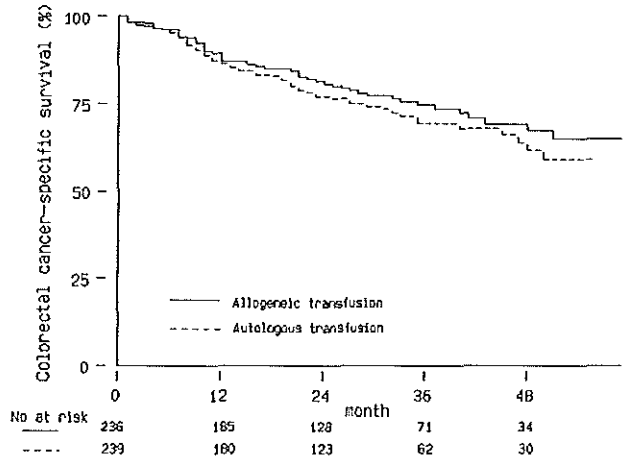


Figure 2. Colorectal cancer-specific survival in all 475 patients with colorectal cancer. The colorectal cancer-specific survival at four years was 67 percent in the allogeneic transfusion group and 62 percent in the autologous transfusion group (P=0.39). The 95 percent confidence interval for the difference between these percentages (autologous minus allogeneic) ranged from -18 percent to +7 percent.

Colorectal Cancer-Specific Survival:

During the study, 114 patients died of colorectal cancer (53 in the allogeneic-transfusion group and 61 in the autologous-transfusion group). The survival of all 475 eligible patients in the two groups was similar (P=0.39) (Fig. 2).

In addition to the Dukes' stage, the patient's age was also significantly related to colorectal cancer-specific survival - i.e., older patients generally did worse than younger ones. The ratio of the death rate in the autologous-transfusion group to the death rate in the allogeneic-transfusion group, after adjustment for Dukes' stage and age, was 1.1 (95 percent confidence interval, 0.8 to 1.7; P=0.66).

As was the case for disease-free survival, the colorectal cancer-specific survival in the 423 patients who underwent curative surgery was significantly better (adjusted P<0.001) in the patients who did not receive transfusions than in those who did; the survival at four years was 88 percent and 65 percent, respectively. When the patients who received transfusions were subdivided according to the type of transfusions they received (allogeneic, autologous, or both), there were no significant differences (adjusted P=0.60) between the three subgroups; the survival at four years was 64 percent, 68 percent, and 63 percent, respectively. The survival of patients who did not receive transfusions did not differ significantly (adjusted P = 0.86) between the two randomized groups; the survival of these patients at four years was 87 percent in the allogeneic-transfusion group and 88 percent in the autologous-transfusion group. When the analysis was restricted to patients who had one or two transfusions of the same type, the patients receiving autologous transfusions and those receiving alloge-

neic transfusions both had worse survival rates than the patients without transfusions (Table 4).No relation was found between survival and the transfusion of fresh-frozen plasma.

Table 3. Multivariate analysis of factors related to disease-free survival in 423 patients with colorectal cancer who underwent curative surgery.*

Factor	Relative Recurrence Rate	95 percent Confidence Interval	P value
Transfusion group			
Allogeneic**	1	-	-
Autologous	1.1	0.7-1.6	0.74
Dukes' classification			
A**	1	-	-
B	4.0	1.7-9.5	0.002
C	10.8	4.7-25.1	<0.001

* Other factors investigated (age and sex of patients, tumor location, adjacent-organ fixation, degree of differentiation, and tumor size) had no significant additional predictive value with respect to disease-free survival, and the effect of randomization was not significantly influenced by any of these factors.
** Reference category.

Discussion

The results of the retrospective studies of the influence of blood transfusions on survival and the recurrence rate in patients with colorectal cancer are conflicting. The studies in which the prognosis in patients receiving transfusions was poorer may have been biased by the selection of patients[14,15]. One way to avoid such confounding by indication[16] is to conduct a randomized trial comparing the effects of autologous and allogeneic blood transfusions[17,18]. The results of this study indicate that as compared with the use of allogeneic blood, the use of autologous blood either to avoid or to reduce exposure to allogeneic blood neither lowered the recurrence rate nor improved survival in patients who had undergone surgery for colorectal cancer. In accordance with some retrospective studies, the recurrence rate was higher in patients who had received transfusions than in those who had not. For patients given transfusions with allogeneic blood, the increase in the recurrence rate was similar to that in the patients who received only autologous blood. The same applied to the survival of the patients.
Recently, Ness et al.[19] reported no difference in survival in a nonrandomized study in which the effects of allogeneic and autologous blood transfusions were compared in

Table 4. Disease-free survival and colorectal cancer-specific curvival, according to transfusion status and Dukes' classification, in patients with colorectal cancer who underwent curative surgery.

Factor	No. of patients	Disease-free survival at 4 yr	Relative recurrence rate*	Survival at 4 yr	Relative death rate*
No. of transfusions					
0 ***	143	73%	1	88%	1
1 or 2					
Allogeneic	75	59%	2.1**	67%	3.6**
Autologous	102	62%	1.8**°	68%	2.8**°
Dukes' classification					
A ***	78	93%	1	94%	1
B	130	73%	4.3**	85%	1.2
C	112	39%	15.5**	50%	10.8**

* Obtained by multivariate analysis.
** $P<0.05$ for the comparison with the reference category.
*** Reference category.
° Not significantly different ($P>0.60$) from the allogeneic-transfusion group.

patients undergoing radical surgery for prostate cancer. The results of a randomized study comparing both types of transfusion in patients with colorectal cancer were also presented recently[20]. In that study, which included only 120 patients, there were fewer recurrences in the autologous-transfusion group, but on the basis of life-table analysis there were no statistically significant differences between the randomized groups.

An explanation for our findings could be that autologous blood induces the same adverse reactions as allogeneic blood. In animals in which allogeneic transfusions had an adverse effect, no such effect was described for syngeneic blood transfusions.[21,22] Since autologous blood transfusions in humans are comparable to syngeneic transfusions in animals, there is no experimental support for an effect of autologous transfusions on tumor growth. On the other hand, autologous transfusion requires the donation of blood. We have found in rats that the donation of blood can decrease natural-killer-cell activity and stimulate tumor growth.[23-25] Therefore, the patients in the autologous-transfusion group could have had a lower natural-killer-cell activity than the patients in the allogeneic-transfusion group at the time of surgery. However, among the patients who did not receive transfusions, there was no difference in survival between the patients in the autologous-transfusion group, who donated blood, and those in the allogeneic-transfusion group.

The most likely explanation for our findings is that there is no causal relation between blood transfusions and prognosis in patients with colorectal cancer. Thus, the findings in the retrospective studies in which blood transfusion was a determinant of

prognosis were probably due to patient selection. We think it is not the blood transfusions themselves, but rather the circumstances necessitating the transfusions, that are the real determinant of prognosis. The need to give patients transfusions during the perioperative period is obviously determined by a number of factors, such as blood loss, the extent of the tumor and the dissection, and the skill of the surgeon, although we found tumor size not to be a determinant of prognosis in the multivariate analysis. In some retrospective studies the groups receiving transfusions contained more patients with rectal tumors - who have a poorer prognosis than patients with colon cancer - than did the groups not receiving transfusions[26,27]. We found that the higher recurrence rates in both groups receiving transfusions, as compared with the rate in the group receiving no transfusions, were not affected by the location of the tumor. Because of our rules regarding transfusion, there was such a strong relation between blood loss and transfusion that it was impossible to separate these two factors. Other possible reasons for transfusion, such as the extent of the dissection and the skill of the surgeon, are difficult to assess. Although it seems beneficial to operate on patients with colorectal cancer in such a way that blood transfusions are either avoided or minimized, there is no reason, with respect to either cancer recurrence or survival, to use a program of transfusion with autologous blood in patients undergoing surgery for colorectal cancer.

References

1. Blumberg N, Heal JM. Transfusion and host defences against cancer recurrence and infection. Transfusion 1989 ;29: 236-245.
2. Francis DMA. Relationship between blood transfusion and tumor behaviour. Br J Surg 1991; 78: 1420-1428.
3. Gantt CL. Red blood cells for cancer patients. Lancet 1981; 2: 363.
4. Burrows L, Tartter P. Effect of blood transfusions on colonic malignancy recurrence rate. Lancet 1982; 2: 662.
5. Singh SK, Marquet RL, Westbroek DL, Jeekel J. Enhanced growth of artificial tumor metastates following blood transfusion: the effect of erythrocytes, leukocytes, and plasma transfusion. Eur J Cancer Clin Oncol 1987; 23: 1537-1540.
6. Francis DMA, Burren CP, Clunie GJA. Acceleration of B16 melanoma growth in mice after blood transfusion. Surgery 1987; 102: 485-492.
7. Tartter PI, Francis DMA. Blood transfusion and tumor growth. Transplant Proc 1988; 20:1108-1111.
8. Jeekel J, Eggermont AMM, Heysteck GA, Marquet RL. Inhibition of tumor growth by blood transfusions in the rat. Eur Surg Res 1982; 14: 124-145 (abstract).
9. Brunson ME, Alexander JW. Mechanisms of transfusion-induced immunosuppression. Transfusion 1990; 30: 651-658.
10. Hoynck van Papendrecht MAW, Hop W, Langenhorst BLAM, Kothe FC, Marquet RL, Jeekel J. Feasability of a predeposit autologous blood donation program in colorectal cancer patients: results from a randomized clinical study. Vox Sang 1992; 62: 102-107.
11. Council on Scientic Affairs. Autologous blood transfusions. JAMA 1986; 256: 2378-2380.
12. Turnbull RB Jr, Kyle K, Watson FR, Spratt J. Cancer of the colon: the influence of the no-touch isolation technic on survival rates. Ann Surg 1967; 166: 420-427.
13. Cox DR. Regression models and life-tables. J Roy Stat Soc 1982; B34: 187-202.
14. Bentzen SM, Balslev I, Pedersen M, et al. Blood transfusion and prognosis in Dukes' B and C colorectal cancer. Eur J Cancer 1990; 26:457-463.

15. Nathanson SD, Tilley BC, Schultz L, Smith RF. Perioperative allogeneic blood transfusions: survival in patients with resected carcinomas of the colon and rectum. Arch Surg 1985; 120: 734-738.
16. Miettinen OS. The need for randomization in the study of intended effects. Stat Med 1983; 2: 267-271.
17. van Lawich van Pabst WP, Langenhorst BLAM, Mulder PGH, Marquet RL, Jeekel J. Effect of perioperative blood loss and perioperative blood transfusions on colorectal cancer survival. Eur J Cancer Clin Oncol 1988; 24: 741-747.
18. Taylor RMR, Parrott NR. Red alert. Br J Surg 1988; 75: 1049-1050.
19. Ness PM, Walsh PC, Zahurak M, Baldwin ML, Piantadosi S. Prostate cancer recurrence in radical surgery patients receiving autologous or homologous blood. Transfusion 1992; 32: 31-36.
20. Heiss MM, Jauch KW, Delanoff CH, Mempel W, Schildberg FW. Blood transfusion modulated tumor recurrence: a randomized study of autologous versus homologous blood transfusion in colorectal cancer. Proc Am Soc Clin Oncol 1992; 11: 172 (abstract).
21. Singh SK, Marquet RL, de Bruin RWF, Westbroek DL, Jeekel J. Promotion of tumor growth by blood transfusions. Transplant Proc 1987; 19: 1473-1474.
22. Waymack JP, Chance WT. Effect of blood transfusions on immune function. IV. Effect on tumor growth. J Surg Oncol 1988; 39: 159-164.
23. Singh SK, Marquet RL, de Bruin RWF, Hop WCJ, Westbroek DL, Jeekel J. Consequences of blood loss on growth of artificial metastases. Br J Surg 1988; 75: 377-379.
24. Hoynck van Papendrecht MAW, Busch ORC, Jeekel J, Marquet RL. The influence of blood loss on tumour growth: effect and mechanism in an experimental model. Neth J Surg 1991; 43: 85-88.
25. Herberman RBB, Ortaldo JR. Natural killer cells: their role in defenses against disease. Science 1981; 214: 24-30.
26. Weiden PL, Bean MA, Schultz P. Perioperative blood transfusion does not increase the risk of colorectal cencer recurrence. Cancer 1987; 60: 870-874.
27. Frankish PD, McNee RK, Alley PG, Woodfield DG. Relation between cancer of the colon and blood transfusion. BMJ 1985; 290: 1827.

CHAPTER 5

QUALITY OF PALLIATION AND POSSIBLE BENEFIT OF EXTRA-ANATOMIC RECONSTRUCTION IN RECURRENT DYSPHAGIA AFTER RESECTION OF CARCINOMA OF THE ESOPHAGUS

This chapter was published in the Journal of the American College of Surgeons 1994; 179: 705-713 by J.J.B. van Lanschot, W.C.J. Hop, M.H.J. Voormolen, R.A.J. van Deelen, J.G.A.M. Blomjous and H.W. Tilanus

Introduction

After operative resection for carcinoma of the esophagus, even if performed with curative intent, most patients die either as a result of locoregional tumor recurrence or distant metastatic disease, or both[1,2]. For these patients, the main advantage of the operation is the relief of severe dysphagia. Unfortunately, in a number of patients, dysphagia recurs early after the operation and is mainly the result of benign fibrotic stenosis at the site of the anastomosis. In most patients, this can be treated successfully by repeated endoscopic dilatation[3]. When dysphagia recurs at a relatively long time after the operation, it almost invariably indicates locoregional tumor recurrence.
In the case of a thoracoabdominal esophagectomy with an intrathoracic anastomosis, the neoesophagus can only be placed in the prevertebral position, that is, in the bed of the tumor[4,5]. During the last few years, many institutions have changed from a thoracoabdominal to a cervicothoracoabdominal procedure with a cervical anastomosis[6-8]. Using this technique, there is the choice of different routes for reconstruction. The extra-anatomic retrosternal route is easy to create and has the advantage that intrathoracic prevertebral tumor recurrence cannot invade the neoesophagus.
This study was done to determine the incidence of secondary dysphagia because of locoregional tumor recurrence after esophageal resection and prevertebral gastric tube reconstruction, and to determine in which patients the recurrent dysphagia could probably have been prevented by using an extra-anatomic reconstruction route.

Patients and methods

In a retrospective study, the records were reviewed of all patients who underwent resection of a malignant tumor of the esophagus or the gastroesophageal junction between 1983 and 1989 in the Department of General Surgery of the Erasmus University Hospital, Rotterdam, The Netherlands. Only patients who had transhiatal or transthoracic esophagectomy and proximal gastrectomy followed by prevertebral gastric tube reconstruction were included. Patients who had reconstruction by colonic interposition or by retrosternal gastric tube were excluded. No defined lymph node dissection was performed. Only lymph nodes immediately adjacent to the primary tumor and lymph nodes along the lesser curvature were resected. Lymph nodes at the base of the left gastric artery were systematically investigated.
All patients had an extensive preoperative examination, comprising esophagogastroscopy with histologic biopsy, indirect laryngoscopy, ultrasound examination of the abdomen and occasionally ultrasound of the neck, and computed tomograpy (CT)

scanning of the thorax and the abdomen. If the tumor was located in the proximal part of the intrathoracic esophagus, bronchoscopy was performed to exclude ingrowth of tumor in the trachea or bronchi.

Patients were only operated upon with curative intent, that is, in the absence of distant metastatic disease and in the absence of signs of local irresectability. Positive nodes either at the celiac trunk or in the neck, or both, were considered to be distant metastases and therefore a contraindication for operative therapy.

During the first part of the study period, all patients underwent preoperative radiotherapy unless the tumor was limited to the cardia. Preoperative radiotherapy was given to tumor dose of 40 Gy (fractions of 2 Gy in four weeks). In 1987, preoperative radiotherapy was abandoned.

Squamous cell carcinomas and adenocarcinomas of the esophagus were classified according to the pTNM-criteria for carcinoma of the esophagus established by the International Union Against Cancer (IUAC) in 1987. Classification of adenocarcinomas of the cardia was done according to the criteria for carcinoma of the stomach (Table 1)[9].

Table 1. The pTNM-classification for tumors of the esophagus and of the cardia according to the 1987 criteria of the international union against cancer.

Carcinoma of the Esophagus			
T-primary tumor			
T_{is}	Carcinoma *in situ*		
T_1	Tumor invades lamina propria or submucosa		
T_2	Tumor invades muscularis propria		
T_3	Tumor invades adventitia		
T_4	Tumor invades adjacent structures		
N-regional lymph nodes			
N_0	No regional lymph node metastasis		
N_1	Regional lymph node metastasis		
M-distant metastasis			
M_0	No distant metastasis		
M_1	Distant metastasis (including cervical and celiac nodes)		
Stage grouping			
Stage 0	T_{is}	N_0	M_0
Stage I	T_1	N_0	M_0
Stage IIA	T_2	N_0	M_0
	T_3	N_0	M_0
Stage IIB	T_1	N_1	M_0
	T_2	Any N	M_0
Stage III	T_3	N_1	M_0
	T_4	Any N	M_0
Stage IV	Any T	Any N	M_1

Table 1 (continued).

Carcinoma of the Cardia

T-primary tumor

T_{is}	Carcinoma *in situ:* intraepithelial tumor without invasion of the lamina propria
T_1	Tumor invades lamina propria or submucosa
T_2	Tumor invades muscularis propria or subserosa
T_3	Tumor penetrates the serosa (visceral peritoneum) without invasion of adjacent structures

N-regional lymph nodes

N_0	No regional lymph nodes
N_1	Metastasis in perigastric lymph node(s) within 3 cm of the edge of the primary tumor
N_2	Metastasis in perigastric lymph node(s) more than 3 cm from the edge of the primary tumor or in lymph nodes along the left gastric, common hepatic, spenic or celiac arteries

M-distant metastasis

M_0	No distant metastasis
M_1	Distant metastasis

Stage grouping

Stage 0	T_{is}	N_0	M_0
Stage IA	T_1	N_0	M_0
Stage IB	T_1	N_1	M_0
	T_2	N_0	M_0
Stage II	T_1	N_2	M_0
	T_2	N_1	M_0
	T_3	N_0	M_0
Stage IIIA	T_2	N_2	M_0
	T_3	N_1	M_0
	T_4	N_0	M_0
Stage IIIB	T_3	N_2	M_0
	T_4	N_1	M_0
Stage IV	T_4	N_2	M_0
	Any T	Any N	M_1

Radicality of the resection was scored as R0 in cases of a microscopically radical resection, as R1 in cases of microscopic tumor residue, and as R2 in cases of macroscopic tumor residue.

Frozen section analysis of the proximal and distal resection margins was not routinely performed. Patients had follow-up evaluation to January 1, 1991, or until death. During the follow-up period, special attention was given to evidence of locoregional

tumor recurrence.
For the pTNM-classification of the primary tumor, positive cervical nodes were scored as distant metastases according to the criteria of the IUAC. During the follow-up period, the cervical metastases were classified as locoregional recurrence because the symptoms were frequently similar to those of intrathoracic tumor.
In cases of secondary dysphagia as a result of tumor recurrence, the exact localization of the recurrence (cervical or intrathoracic) was assessed and whether or not the recurrence originated intramurally near the anastomosis (so-called central recurrence) or invaded the neoesophagus from outside (peripheral recurrence) was determined. Only in cases of peripheral, intrathoracic tumor recurrence causing upper gastrointestinal symptoms (especially dysphagia) for at least one month was it concluded that the extra-anatomic route would have been beneficial.
The overall survival and local recurrence-free survival rates were calculated according to the Kaplan-Meier method. Comparisons were done with the log-rank test. Cox regression was used to evaluate the prognostic value of various factors simultaneously with respect to locoregional recurrence ($p=0.05$ [two-sided] was considered the limit of significance)[10]. An adapted Kaplan-Meier method was used to determine the absolute risk of secondary dysphagia as a result of locoregional tumor recurrence. As this method is not well known, details are given in the Appendix.

Results

Two hundred nine patients fulfilled the selection criteria. The age of the patients at the time of the operation ranged from 31 to 84 years, with a mean age of 61.3 years. The male to female ration was 5:2. Of the 209 patients, 143 received preoperative radiotherapy. Ninety-five patients had squamous-cell carcinoma of the esophagus (group 1) and 57 patients had adenocarcinoma of the esophagus, almost invariably arising in Barrett mucosa (group 2). In 57 patients, the tumor originated from the cardia (group 3). The main tumor characteristics are summarized in Table 2.

Twelve patients died postoperatively, eight patients with squamous-cell carcinoma, three patients with adenocarcinoma of the esophagus, and one patient with adenocarcinoma of the cardia, resulting in an in-hospital mortality rate of 8, 5 and 2 percent, respectively. Follow-up evaluation was complete in 208 patients (one patient left the country three years after resection, without signs of recurrent disease). The mean total follow-up period postoperatively was 2.2 years (range of 0.1 month to 7.9 years). The mean follow-up period of the survivors was 4.9 years (range of 2.3 to 7.9 years). The overall five-year survival rate for the three groups is illustrated in Figure 1.
During the follow-up period, 73 patients showed various symptoms of locoregional tumor recurrence, cervical nodes included (Table 3). The diagnostic procedures, which were performed to confirm and to localize the recurrent tumor mass, are summarized in Table 4. In 52 (71 percent) of the 73 patients, locoregional tumor recurrence was confirmed histocytologically.
The results of the multivariate analysis of various risk factors with respect to locoregional recurrence are given in Table 5. The most important prognostic factors were the N1 and M1 category (positive celiac nodes). No significant additional predictive value was found for deep local invasion (T3.4 versus T1.2) or macroscopic tumor residue (R2 versus R0.1), although both factors significantly correlated with locoregional

Table 2. Main tumor characteristics of 209 patients who had operative therapy for a malignant tumor of the esophagus or gastroesophageal junction.

Group 1: Squamous cell carcinoma, $n=95$
Group 2: Adenocarcinoma of the esophagus, $n=57$
Group 3: Adenocarcinoma of the cardia, $n=57$

Tumor characteristics	Group 1	Group 2	Group 3
pTNM stage			
1	14	8	11
2	66	33	19
3	9	9	25
4	6	7	2
Differentiation			
Good	9	3	2
Moderate	50	36	21
Poor	31	17	32
Undifferentiated	5	1	2
Microscopic tumor residue (R_1)			
Only proximal margin positive	0	1	2
Only distal margin positive	0	0	3
Proximal and distal margins positive	0	0	1
Only tangential margin positive	1	2	3
Macroscopic tumor residue (R_2)	16	6	1

recurrence when evaluated univariately. Locoregional recurrence-free survival according to pT, pN, pM, and R category for the three tumor groups combined are illustrated in Figure 2.

No significant prognostic value was found for the tumor type (squamous, Barrett, cardia). The same applied to age and gender of the patients, location of the primary tumor (upper, middle or lower third), degree of differentiation, and preoperative irradiation.

Locoregional tumor recurrence caused upper gastrointestinal symptoms in 46 patients (22 percent). For these patients, the operation did not give definite palliation with respect to normal food intake (Tables 3 and 6). The median time between operation and the occurrence of recurrent dysphagia as a result of locoregional metastases was 24 months (range of two to 42 months). The absolute risk of malignant secondary

Table 3. Locoregional tumor recurrence according to symptoms of 209 patients who had operative therapy for a malignant tumor of the esophagus or gastroesophageal junction.

Group 1: Squamous cell carcinoma, *n=95*
Group 2: Adenocarcinoma of the esophagus, *n=57*
Group 3: Adenocarcinoma of the cardia, *n=57*

Recurrence and symptoms	Group 1	Group 2	Group 3	Total
Locoregional recurrence	34 (36)	22 (39)	17 (30)	73 (35)
Symptoms* related to upper gastrointestinal tract				
Dysphagia/vomiting	21 (22)	11 (19)	14 (25)	46 (22)
Aspiration due to fistula	2 (2)	1 (2)		3 (1)
Hematemesis	1 (1)	1 (2)		2 (1)
Other symptoms*				
Pain	14 (15)	12 (21)	12 (21)	38 (18)
Respiratory obstruction	14 (15)	5 (9)	7 (12)	26 (12)
Hemoptysis	3 (3)	1 (2)	1 (2)	5 (2)
Hoarseness	13 (14)	7 (12)	1 (2)	21 (10)

* Patients can have more than one symptom. Numbers in parentheses are percentages.

Table 4. Number of additional diagnostic procedures to confirm and to localize the recurrent tumor.

	All patients* *n=73*	Group 1 *n=34*	Group 2 *n=22*	Group 3 *n=17*
None	3	2	1	0
Chest roentgenogram	33	24	7	2
Endoscopy	48	20	13	15
Ultrasound	28	18	2	8
CT-scan	25	15	6	4
Histology, cytology	52	20	16	16

* Many patients had more than one test.
CT: computed tomography.

dysphagia amounted to 18 and 22 percent at two and five years, respectively, for the three tumor groups combined (Figure 3). By the conclusion of the study period, all patients with recurrent malignant dysphagia had died. Recurrent dysphagia lasted an average 5.3 months (Range of 0.3 to 21.5 months), before the patients died (Table 6).

Table 5. Multivariate analysis of locoregional tumor recurrence in relation with the histopathologic factors of 209 patients who had operative therapy for a malignant tumor of the esophagus or gastroesophageal junction.

Group 1: Squamous cell carcinoma
Group 2: Adenocarcinoma of the esophagus
Group 3: Adenocarcinoma of the cardia

	Total, *n=209*	Recurrence at 5 yr (percent)*	Relative recurrence rate°	p value**
Tumor type				
Group 1	95	44	1.0	-
Group 2	57	53	0.9	0.79
Group 3	57	53	0.8	0.42
pT-category				
$T_{1,2}$	78	37	1.0	-
$T_{3,4}$	131	51	1.5	0.15
pN-category				
N_0	146	39	1.0	-
N_1	63	74	2.3	0.002
pM-category				
M_0	195	45	1.0	-
M_1	14	>90	5.2	<0.001
R-category				
$R_{0,1}$	186	45	1.0	-
R_2	23	69	1.8	0.10

* Actuarial percentage (univariate).
** For comparison with the reference category.
° 1.0 denotes the reference category.

In 27 of the 46 patients with malignant secondary dysphagia, the recurrent tumor mass was localized within the chest and invaded the neoesophagus from outside. Only for this subgroup of patients with intrathoracic peripheral tumor recurrence could the upper gastrointestinal symptoms have been prevented by using an extra-anatomic reconstruction route. The mean duration of this probable advantage was 4.4 months (range of one to 15.6 months) (Table 6). Analysis of the various risk factors related to this specific type of recurrence showed similar results, as described for the whole group of patients with locoregional recurrence (Figure 4). Six of the 27 patients with a symptomatic, peripheral, intrathoracic recurrence had an R2 resection. In the univariate analysis, the most important prognostic factors were the N1 and M1 category and the presence of macroscopic tumor residue (R2 compared with R0.1).

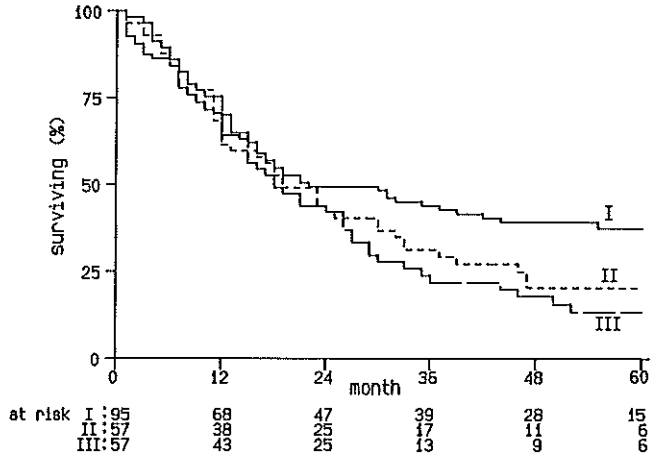


Figure 1. The overall survival rate of 209 patients who had operative therapy for a malignant tumor of the esophagus or gastroesophageal junction. Group I, squamous cell carcinoma; group II, adenocarcinoma of the esophagus, and group III, adenocarcinoma of the cardia. Number of patients at risk are indicated at the bottom. The difference between group I and group III is significant (P=0.02).

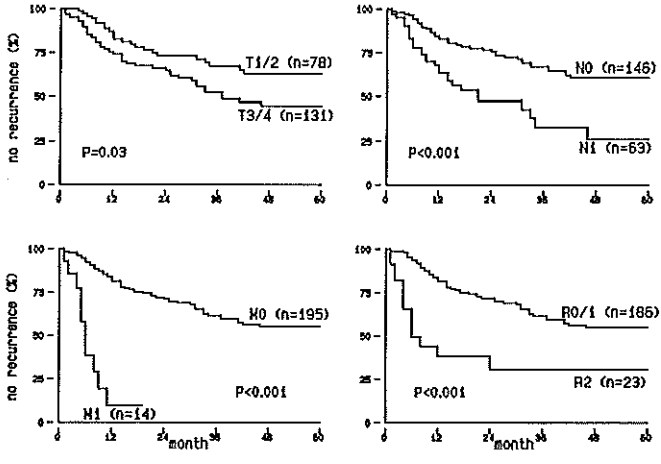


Figure 2. Locoregional recurrence-free survival rate for the three tumor groups combined (n=209) according to pT-category, pN-category, pM category and R-category. P-values refer to univariate comparisons.

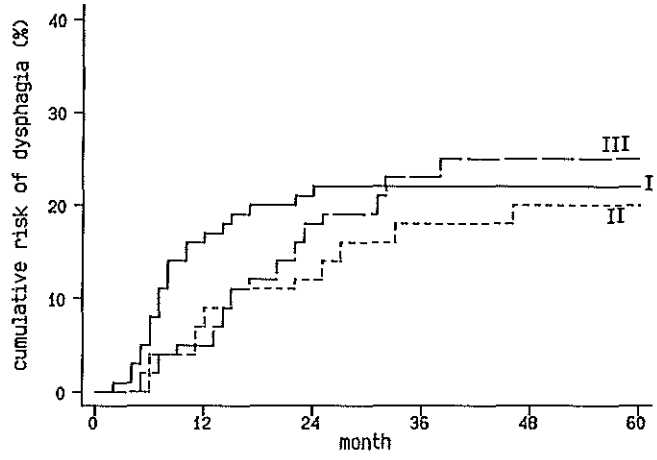


Figure 3. Absolute risk of secondary dysphagia as a result of locoregional tumor recurrence in 209 patients who had operative therapy for a malignant tumor of the esophagus or gastroesophageal junction. Group I, squamous cell carcinoma; group II, adenocarcinoma of the esophagus, and group III, adenocarcinoma of the cardia.

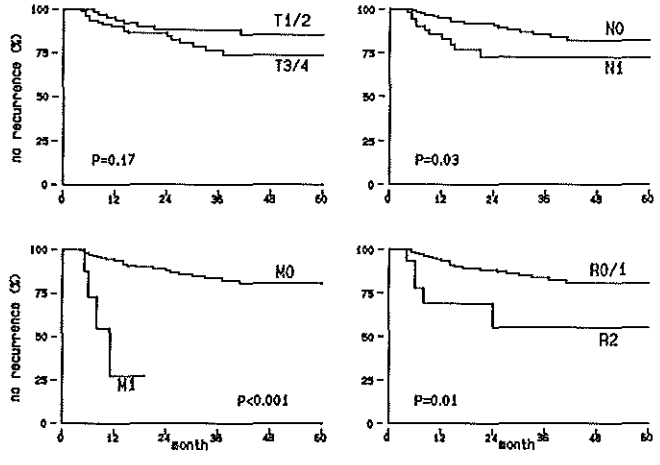


Figure 4. Peripheral, intrathoracic recurrence-free survival rate for the three tumor groups combined (n=209) according to pT-category, pN-category, pM-category and R-category. P-values refer to univariate comparisons.

Table 6. Locoregional tumor recurrence in relation to a possible benefit of an extra-antomic reconstruction route in 209 patients who had operative therapy for a malignant tumor of the esophagus or gastroesophageal junction

Group 1: Squamous cell carcinoma, $n=95$
Group 2: Adenocarcinoma of the esophagus, $n=57$
Group 3: Adenocarcinoma of the cardia, $n=57$

(Duration from onset up to death)

	Group 1	Group 2	Group 3	Total
Locoregional recurrence	34 (36)	22 (39)	17 (30)	73 (35)
Locoregional recurrence plus upper GI symptoms	21 (22)	11 (19)	14 (25)	46 (22)
Mean duration upper GI symptoms*	5.4	4.8	5.6	5.3
Median and range duration*	3.1 0.3-21.5	4.8 1.2-15.4	3.7 1.0-15.6	3.5 0.3-21.5
Benefit extra-anatomic reconstruction	11 (12)	9 (16)	7 (12)	27 (13)
Mean duration extra-anatomic benefit*	4.4	3.7	6.0	4.4
Median and range duration benefit*	2.8 1.0-9.5	1.2 1.0-11.4	13.0 1.0-15.6	2.7 1.0-15.6

Numbers in parentheses are percentages.
GI: gastrointestinal tract.
* Months.

Discussion

After operative therapy for carcinoma of the esophagus or the gastroesophageal junction with curative intent, about one-third of the patients had locoregional tumor recurrence. In most patients, this resulted in disabling upper gastrointestinal tract symptoms, especially in recurrent dysphagia. For these patients, operative therapy provided neither definite cure nor permanent palliation of dysphagia. When evaluating the quality of palliation by operative therapy compared with other treatment modalities for carcinoma of the esophagus, one should not only consider the in-hospital morbidity and mortality rates, but also the early and the late recurrence of dysphagia[11]. Early dysphagia is almost always caused by benign stricture formation at the site

of the anastomosis and can effectively be treated by endoscopic dilatation[12].
If dysphagia recurs relatively late postoperatively, it almost invariably indicates tumor recurrence. Forty-six patients (22 percent) had secondary dysphagia as a result of locoregional tumor recurrence (positive cervical nodes included). The operative therapy for these patients must be considered as failures, both with respect to definitive cure and to permanent palliation. Given these circumstances, additional palliative measures, such as radiotherapy and pertubation, are indicated, depending on the general condition of the patient, the localization and the extension of the recurrent tumor, and the presence of distant metastases. In 27 (59 percent) of these 46 patients, the recurrent tumor mass was localized within the chest and invaded the neoesophagus from outside. For this subgroup of patients, an extra-anatomic reconstruction route would probably have prevented the upper gastrointestinal tract symptoms. This advantage would have lasted on an average for 4.4 months, before the patients died.
The risk of locoregional tumor recurrence is clearly affected by the regional dissemination and the radicality of the resection. The presence of positive lymph nodes (N1), especially at the celiac trunk (M1) and a macroscopically nonradical R2 resection, seem to be important risk factors. Although all patients had lymph nodes present in the resected specimens, the number of nodes was limited. This might explain the relatively high percentage of patients with N0 disease when compared with patients with an extended lymphadenectomy.
The minimal follow-up period of the patients was two years. Although a substantial number of patients were still alive without disease by the end of the study period, and therefore still at risk for tumor recurrence, we know from earlier studies that most tumor-related events occur within two years postoperatively[2]. This can explain why the absolute risk of having malignant secondary dysphagia hardly rises after the second year (Figure 3).

What is the price to be paid for the advantage of an extra-anatomic reconstruction? The closure of the hiatus and the creation of a retrosternal tunnel are simple and take only a few minutes. A rare but serious perioperative complication is retrosternal (venous) bleeding, for which a sternotomy is sometimes necessary to control the hemorrhage. If the whole stomach is brought up to the neck, it can be necessary to remove a portion of the manubrium and the medial portion of the left clavicle to avoid compression of the gastric conduit. If a narrow gastric tube is used (2 to 3 cm), this is hardly ever indicated. The retrosternal route is 2 to 3 cm longer than the prevertebral route, which might induce traction and relative ischemia at the site of the anastomosis[13,14] . This might explain the increased risk of anatomic leakage that has been reported for palliative retrosternal gastric bypass procedures of the excluded esophagus[15]. However, it is unclear from these uncontrolled data to what extent the high leakage rate must be attributed to the poor general condition of these patients with advanced malignancy.
No data are available on the functional aspects of a gastric tube in an extra-anatomic position. In case of a retrosternal reconstruction, the neoesophagus follows the manubrium in a curved course, which might hamper the normal passage of food and liquids. Furthermore, endoscopic dilatation of a benign stricture might be more complicated because of this curved course. Institutions at which the extra-anatomic route is always used after "curative" resection, do not report a high incidence of

technical complications, functional complications, or both[16].

APPENDIX

The Kaplan-Meier method (Kaplan EL, P Meier: Nonparametric estimation from incomplete observations. J Am Statist Assoc 1958, 53: 457-481) has proved to be useful for providing information about the pattern of events, for example death or tumor recurrence, along time. Essentially, the method adjusts for the duration of follow-up, which, in nearly all cases, varies from patient to patient. The Kaplan-Meier method gives an estimate of the risk of an event within a certain period for an event bound to occur in each patient, provided that the follow-up period is sufficiently long, for example, death irrespective of cause. Such an interpretation in term of risks is, in general, not possible for events that do not invariably occur in each patient, for instance, local tumor recurrence (Pepe MS, M Mori: Kaplan-Meier, marginal or conditional probability in summarizing competing risk failure time data? Stat Med 1993; 12: 737-751). Patients who die, possibly as the result of distant metastases or intercurrent disease before a local tumor recurrence becomes manifest, are handled by the Kaplan-Meier method as if they left the study at the moment of that death. Thereafter, however, these patients are imaginary still considered to be at risk for tumor recurrence. This is based on the assumption that if such deaths could be avoided, such patients would have the same future risk of having a recurrence. Because it is impossible to prevent death without tumor recurrence, the Kaplan-Meier curve will not depict the true risk, along time, for patients to develop a local recurrence, but will give an over-estimate of that risk. This is especially the case in studies on patients with a high mortality from other causes.

This phenomenon can be demonstrated by a hypothetical example. Suppose that 75 of a group of 100 patients die within one year without local tumor recurrence, and the remaining 25 patients have local recurrence during the second year of follow-up. The Kaplan-Meier method will then show a percentage of 100 without local recurrence in the first year of follow-up study. This percentage will decrease to zero in the second year of follow-up study, suggesting that the "risk" of local tumor recurrence for the total group of patients equals 100 percent. This 100 percent risk, however, applies only to the subgroup of patients who survive the first year. Because only 25 of the 100 patients actually have local recurrence during their life, the risk for the total group is 25 of 100 (25 percent) instead of 100 percent. The discrepancy occurs because the patients who died without local recurrence should no longer be considered as being at risk for recurrence.

A method is described to estimate the true (absolute) risk of an event to occur during a certain period of time (T months) after entering the study. The event to be considered we derive from this article and concerns secondary dysphagia as a result of locoregional tumor recurrence in patients operated upon for carcinoma of the esophagus. The method is presented in two steps. The probability needs to be estimated that patients have secondary dysphagia in a particular month. This probability equals the probability that patients are alive in that same month without a previous dysphagia, multiplied with the probability that patients have secondary dysphagia in the mentioned month. In this multiplication, the first probability is estimated by the standard Kaplan-Meier estimate, the second is estimated by the

proportion of patients who have secondary dysphagia among those who are still at risk in that particular month. That is, with exclusion of the patients who experienced the event before, had died or had left the study for other reasons. The first step has to be repeated for each month preceding the time T.
The second step concerns addition of all the calculated probabilities for the months preceding the month T. The sum thus obtained gives the estimate of the risk of secondary dysphagia before the time T. If all the patients have follow-up evaluation at least to the time T, in case they did not die before that time, this method simply reduces to the calculation of the proportion of patients who did experience the event before the time T, with the total number of the group as denominator.
According to this method, the risk for the total group of patients, reported in the present study to have secondary dysphagia within two years (T=24 months) or five years (T=60 months) postoperatively, seemed to be 18 and 22 percent, respectively. In contrast, with the original Kaplan-Meier method, these percentages were calculated as 23 percent within two years and 32 percent within five years, thereby considerable over-estimating the absolute risks. It is likely that some of the patients who died during the early follow-up period "escaped" from dysphagia. This has obviously influenced the risk of dysphagia itself, but should not result in adjustments of the estimate of the risk.

References

1. Eeftinck Schattenkerk M, H Obertop, HJ Mud, et al. Survival after resection for carcinoma of the oesophagus. Br J Surg 1987;74: 165-158.
2. Muller JM, H Erasmi, M Stelzner, at al. Surgical therapy of oesophageal carcinoma. Br J Surg 1990;77: 845-857.
3. Pierie JPEN, PW De Graaf, H Poen, et al. Incidence and management of benign anastomotic stricture after cervical oesophagogastrostomy. Br J Surg 1993; 80: 471-474.
4. Ong GB, KH Kwong. The Lweis-Tanner operation for cancer of the oesophagus. J R Coll Surg Edinb 1969; 14: 3-9.
5. Tam PC, HC Cheung, L Ma, et al. Local recurrences after subtotal esophagectomy for squamous cell carcinoma. Ann Surg 1987; 205: 189-194.
6. Orringer MB. Transhiatel esophagectomy without thoracotomy for carcinoma of the thoracic esophagus. Ann Surg 1984; 200: 282-288.
7. Szentpetery S, T Wolfgang, RR Lower. Pullthrough esophagectomy without thoracotomy for esophageal carcinoma. Ann Thor Surg 1979; 27: 399-403.
8. Finley RJ, M Grace, JH Duff. Esophagectomy without thoracotomy of the cardia and lower part of the esophagus. Surg Gyneacol Obstet 1985; 160: 49-56.
9. International Union Against Cancer: TNM classification of malignant tumours . 4th fully revised edition. Edited by P Hermanek and LH Sobin. Springer Verlag, Heidelberg, 1987.
10. Cox DR. Regression models and life tables. J Roy Stat Soc 1972; B34: 187-202.
11. Bown SG. Palliation of malignant dysphagia: surgery, radiotherapy,laser, intubation alone or in combination? Gut 1991; 32: 841-844.
12. Tilanus HW, WCJ Hop, BLAM Langenhorst, JJB van Lanschot. Esophagectomy with or without thoracotomy: is there any difference? J Thorac Cardiovasc Surg 1993; 105: 898-903.
13. Orringer MB, H Sloan. Substernal gastric bypass of the excluded thoracic esophagus for palliation of esophageal carcinoma. J Thorac Cardiovasc Surg 1975; 70: 836-851.
14. Ngan SYK, J Wong. Lengths of different routes for esophageal replacement. J Thorac Cardiovasc Surg 1986; 91: 790-792.

15. Orringer MB. Substernal gastric bypass of the excluded esophagus- results of an ill-advised operation. Surgery 1984; 96: 467-470.
16. Akiyama H, M Tsurumaru, T Kawamura, Y Ono. Principles of surgical treatment for carcinoma of the esophagus. Ann Surg 1981; 194: 438-446.

CHAPTER 6

EVALUATION OF CHANGES IN PROGNOSTIC STATUS

This chapter was published in shorter form in Computers in Biology and Medicine 1989; 19: 181-188 by W.C.J. Hop and H.R. van Buuren

6.1 Introduction

The methodology of fixed factors, that is factors which have a fixed level throughout the study and which were determined at the time of entry into the study is reviewed in Chapter 2. Basic elements in the evaluation of such factors is the grouping of patients according to the levels of the factors, and comparing the groups thus obtained using Kaplan-Meier survival curves and significance tests. Another approach in the evaluation of such fixed factors would be to model the hazard rate, e.g. by using Cox-regression [Cox, 1972]. These approaches often lead investigators erroneously to apply the same techniques to evaluate factors which are time-dependent, i.e. factors which manifest themselves during follow-up. As an example one may consider the prognostic impact of a recurrence of disease regarding the ultimate survival, or some other event observed during follow-up. By grouping the patients according to whether or not the recurrence had occurred, and comparing the two groups thus obtained regarding their survival taken from entry into study, a severely biased outcome may result. This bias arises from the fact that the length of survival itself will influence the chance of a patient to be classified into one group or the other. A late recurrence necessarily implies a long survival from entry. A preferable strategy is to investigate the survival pattern following the event studied. But then the problem is how the contrast to such a survival curve is to be defined. The survival curve of the total patient group is often used for comparison, but one can easily construct examples in which the patterns of occurrence of the event and death are such that misleading conclusions about the effect of the event may be reached. In the field of clinical oncology, where responders to chemotherapy often are compared with nonresponders regarding survival, such so called response-time bias is well recognized [Anderson et al, 1985]. Each patient starts in the non-response category and can only move to the responder status after sufficient time. Patients with an early death thus will not have the opportunity to enter the responder group. If it is required to compare survival according to response category, proper adjustments for the response time are needed. Of course, such analyses can not discriminate between the situation in which survival is affected by the response and the one in which response is only a marker of better prognosis. Also the presence of already known prognostic markers at baseline has to be taken into account [Oye et al, 1984]. Similar problems occur when one wants to correlate retrospectively with survival the amount or duration of chemotherapy received [Redmond et al, 1983].
A partial solution to the problem of response-time bias is provided by the so-called "landmark"-method.

6.2 The Landmark Method

A way to avoid the phenomenon of response-time bias is to select some "landmark" time point [Anderson et al, 1983]. Only patients are considered who survive the chosen landmark time t_L , and these are classified according to the history of the factor, e.g. a response, whose prognostic impact is to be evaluated. The choice of the time-point t_L is arbitrary but may be based on some natural time of clinical significance. Also the number of patients available may give some guidance in the choice of t_L. For the analysis, all methods described in chapter 2 are available to assess the prognostic impact of the factor considered. This landmark method was used by Ettinger and Lagakos (1982) to evaluate the prognostic impact of tumour response in lungcancer patients treated with chemotherapy.
The method is illustrated for patients with inoperable breast carcinoma without evidence of distant metastases who had been treated with radiotherapy [Treurniet-Donker et al, 1980]. It was investigated whether a local tumour recurrence was predictive for a worse survival. For the chosen landmark-time t_L=20 months after start of treatment this is shown in Figure 1.

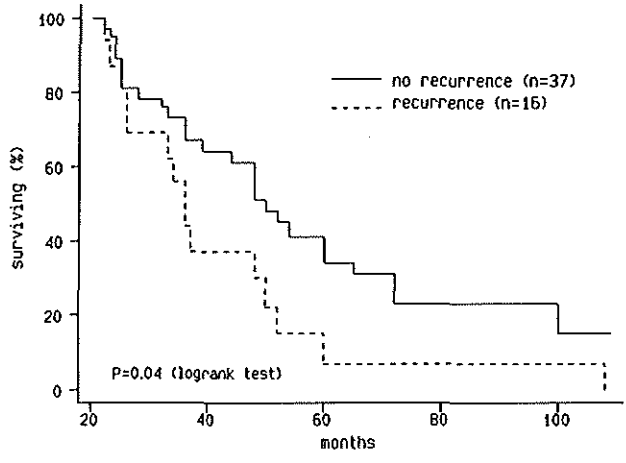


Figure 1. Survival past 20 months after treatment. Patients are grouped according to whether or not the tumour had recurred before 20 months after treatment.

For those patients surviving 20 months it appeared that it was an ominous sign if a local recurrence had developed before that time. Similar findings applied to other choices of t_L.
A major advantage of the landmark method is its conceptual simplicity. A great disadvantage, however, is the neglect of the information about the time behaviour of the factor studied after the landmark time t_L. Also any relation between the prognostic factor and death rates before t_L is not considered. A preferable method therefore is the so-called "transient-state" method.

6.3 The transient state method

Much interest of various biostatisticians in the problems discussed here was raised by the early heart-transplantation studies. Randomized studies in this field are lacking because it is clinically considered not valid to perform these. Therefore problems arise regarding a fully valid evaluation of the efficacy of the procedure. If one considers the group of patients who entered the program, and compares survival of those who actually received a heart transplant with those patients who did not, it is evident that such a comparison will be severely biased [Gail, 1972]. A transplanted patient must have lived long enough while being on the waiting list before a donor heart would become available. No such requirement applies to the non-transplanted patients. There is a real possibility that the instantaneous risk of dying, i.e. the hazard rate for the group as a whole, is decreasing while being on the waiting list due to heterogeneity among the patients, although for individual patients this risk may increase in time. Mantel and Byar [1974] showed that a pseudo control group for transplanted patients could be formed. This is shown in Figure 2 for a number of hypothetical patients. Basically, Mantel and Byar's method is reduced to a comparison of death rates between transplanted patients and their "proper" controls at all possible points in time. Their statistic is very similar to the logrank-statistic, except that group-membership is not fixed but may change along time
[Crowley, 1974].

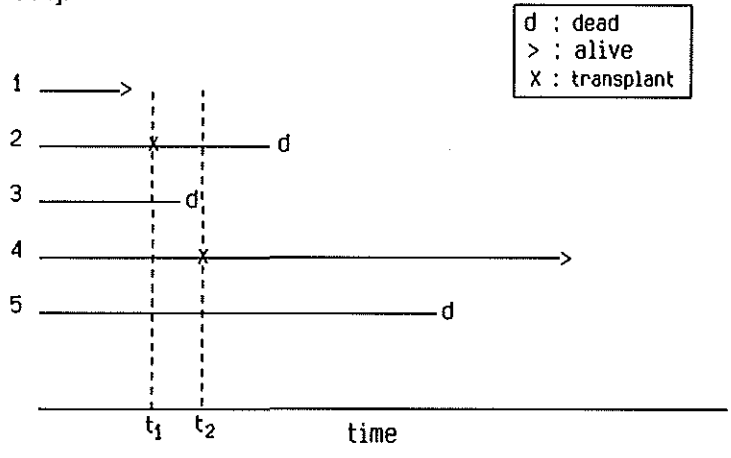


Figure 2. Dynamics of five hypothetical patients. Between time-points t_1 and t_2 patients 3, 4 and 5 serve as controls for patient 2. After t_2 patient 5 acts as control for both patients 2 and 4.

A graphical procedure to accompany the method is described by Simon and Makuch (1984). From some chosen time-point x, which should at least be equal to the smallest observed time at which both prognostic states are present, they suggest to estimate the cumulative hazards $\int_x^t \lambda_0(\tau)d\tau$ and $\int_x^t \lambda_1(\tau)d\tau$. Here $\lambda_0(t)$ and $\lambda_1(t)$ denote, respectively, the death rates while patients are in the prognostic states 0 (e.g. non-transplan-

ted) and 1 (transplanted). Simon and Makuch (1984) further suggest to convert these cumulative hazards into Kaplan-Meier type curves because the latter might be more easy to communicate.
Figure 3 shows the cumulative hazard rates thus obtained for the breast cancer patients as described before. The figure shows that the cumulative hazard associated with recurrences has about twofold increased as compared with patients without a local recurrence. In the following a method is described in which relations between time-dependent factors and survival can be quantified in a more rigorous manner.

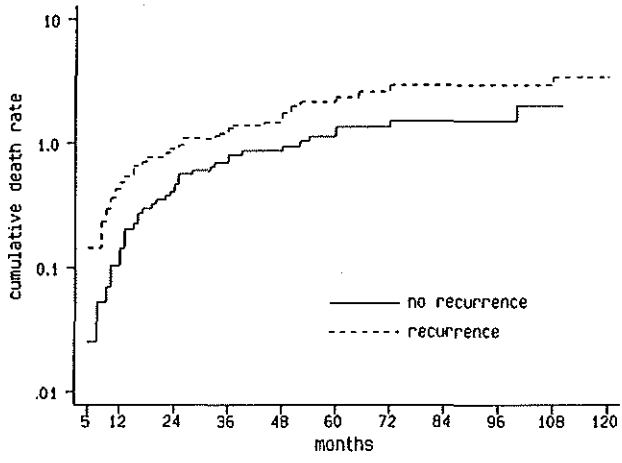


Figure 3. Estimated cumulative death rates according to recurrence status. Note the logarithmically transformed vertical axis.

6.4 Cox-regression with time-dependent covariates

The proportional hazards model (chapter 2, equation 8.1) can be extended to include time-dependent covariates by using

$$\lambda(t;Z(t)) = \lambda_0(t)\, e^{Z(t)\beta} \qquad (6.1)$$

Here Z(t) contains elements whose levels depend on time, e.g. Z(t)=0 and Z(t)=1 denote the two prognostic states as descibed in the previous section. Factors with fixed levels may additionally be represented in Z(t). This model has been used in the evaluation of heart transplantations [Crowley and Hu, 1977]. If the time-dependent covariate Z(t) takes the value 1 as soon as the patient is transplanted, and Z(t)=0 corresponds to the situation before that time, the model then states that at time t after entry into the program, the mortality rate among transplanted patients equals the constant e^{β} times the mortality rate among patients waiting for a transplant. If it appears that e^{β} is smaller than one, i.e. $\beta<0$, such finding would indicate that transplantation is beneficial with respect to survival prognosis. A similar approach was used in an evaluation of allogeneic and autologous bone marrow transplantation in

acute myelogenous leukemia [Hermans et al, 1989]. The procedure is graphically displayed in figure 4.

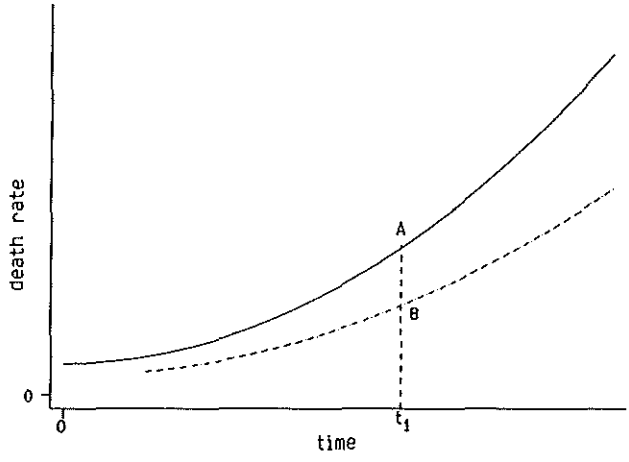


Figure 4. Death rates according to two prognostic states. At time t_1 a transition occurs from A to a more favourable prognostic status B.

Mauger et al (1995) modelled both a short-term and long-term effect associated with the transition to another prognostic state. The effect of such a transition is modelled as

$$\lambda(t;w) = \lambda_0(t)\, e^{(\beta_1+\beta_2 \exp(\beta_3(t-w)))\, I(t-w)} \qquad (6.2)$$

In this formula w denotes the time point at which the transition takes place; I(t-w) equals zero for values of t less than w, and equals 1 if t is equal or greater than w. Before the transition, the death rate equals $\lambda_0(t)$. The death rate immediately after the transition equals $\exp(\beta_1+\beta_2)\lambda_0(t)$, whereas the death rate corresponding to the long-term effect of the transition equals $\exp(\beta_1)\lambda_0(t)$. The parameter β_3 (<0) determines the rate at which the ultimate hazard rate $\exp(\beta_1)\lambda_0(t)$ will be reached. A graphical display of the model is shown in figure 5.
Applying this method for patients who had received a kidney-transplant, Mauger et al (1995) found that such patients initially experienced a somewhat higher death rate as compared to those who remained on dialysis. In the long run, however, their mortality was less.
Kalbfleisch and Prentice (1980) make a distinction between internal time-dependent covariates and external time-dependent covariates. An internal time-dependent covariate is a variable whose level is generated by the individual under study. This level can only be observed as long as the patient is alive and under observation. An external time dependent covariate is one whose level is determined by some external process, i.e. a process not under the control of the individual under study. A donor organ becoming available in the examples above can be considered to be of the latter type. Also the cumulative exposure to particular agents in cohort studies of the inci-

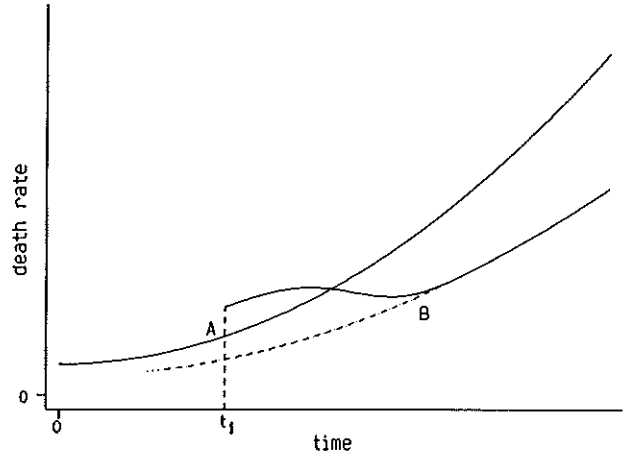


Figure 5. At time t_1 the death rate initially increases from A to a higher value. Ultimately a more favourable death rate is reached at B. The rapidity at which B is reached depends on the third beta-parameter.

dence of particular diseases in occupational epidemiology, can be considered to be of the external type. Factors studied in clinical medicine in most cases are of the internal type. The study of the value of certain biochemical tumour markers in relation to the relapse rate in cancer patients [Gail, 1981] is an example. Another example is the evaluation of the anti-leukemic effect of graft-versus-host disease in patients who have recieved a bone marrow transplant [Prentice et al, 1978]. For such studies the variable Z(t), which may represent not only the value of a particular factor at time t, but may also include aspects about the behaviour of the factor before time t, need to be quantified. An adequate functional choice for Z(t) may be obtained using clinical guidance but is not always easy. For the tumour marker study, Gail (1981) considered some variants of representation of the marker history for an individual up to time t, the most simple one being the level of the tumour marker itself at time t without consideration of previous values. Another aspect of the tumour marker behaviour would be to ascertain whether in the interval from 0 to t one or more elevated values of the marker were observed. In the latter situation the time-dependent covariate for an individual would be a step-function which equals 0 before any elevation is observed. It takes the value 1 after the first occcurrence of an elevated value. In this case Z(t) reduces to a status-indicator as described before. Also the rate of increase of the tumour marker might be a proper choice.
Sometimes a graph of levels of the variable studied after grouping of patients according to the time-point at which the event occurred may be useful. This is demonstrated by the following example. For asthmatic children it was investigated whether the measurement of transcutaneous oxygen tension could be used as a substitute for the measurement of the PD_{20} [Broekhoven et al, 1991]. The PD_{20} denotes the dose of the provocative agent Metacholine which causes a decrease of 20 percent of the forced expiratory flow in one second (FEV_1). Starting from an initial dose of Metacholine, each time the dose was doubled up to the moment that the PD_{20} was

reached. After grouping the patients studied according to the dose at which the PD_{20} was reached, figure 6 is obtained. At the highest dose tested (dosestep 10), two children still did not respond. The figure shows that the oxygen tension behaves relatively stable prior to achieving the required FEV_1 response. At the dose-step at which the PD_{20} was reached, a large decrease of the oxygen tension had taken place. Analysis of the data using the discrete-time version of Cox-regression (there is no continuous time involved here, but discrete dose-steps; chapter 2) showed that the current level of the decrease from baseline of the transcutaneous oxygen tension was to a high degree significantly related to the probability of having reached the PD_{20}.

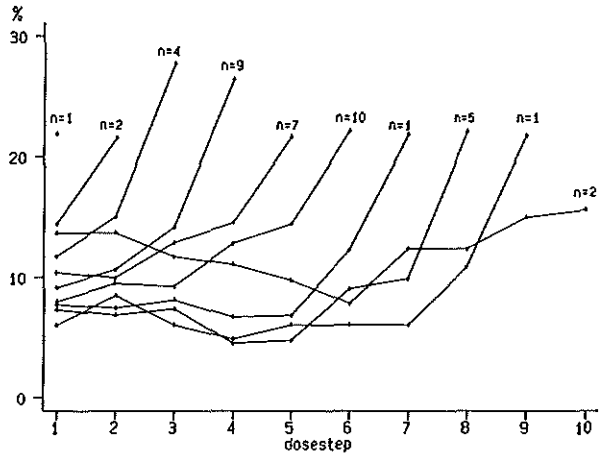


Figure 6. Mean decrease from baseline (percent) of transcutaneous oxygen tension. Patients are grouped according to the dose-step at which the PD_{20} was reached.

To estimate the components of ß, the approximate likelihood (chapter 2, equation 8.4) is generally used, but now with the values of Z in the various risk sets as determined at the time of failure to which the risk sets correspond. For the case with only one status indicator (Z(t) equals 0 or 1), it was shown by Crowley (1974) that the score test of $ß=0$ reduces to the two-groups logrank test with group membership varying with time [Mantel and Byar, 1974]. Special programming using the method of profile likelihood [McCullagh and Nelder, 1990] is required for the case concerning the separate short-term and long-term effects discussed (equation 6.2 above)
Model (6.1) was applied in the case of the irradiated breast cancer patients discussed above. The parameter ß, which represents the effect of a local recurrence on the death rate, was estimated to be 0.58 with a standard error of 0.25. For patients developing a local recurrence the death rate therefore seems to increase by a factor $e^{0.58}$ (=1.8).
The baseline hazard function can be estimated in a similar way as for fixed covariates, with the time-dependent covariates taking the observed values at the different failure times.
Since in clinical practice the covariate values will often not be monitored continuously, but only at fixed intervals, some kind of interpolation will be necessary to obtain the values of the time-dependent covariates at the different failure times for the

various individuals. Simple interpolation rules, such as linear interpolation or carrying forward the last observation, generally will suffice. Some of the goodness-of-fit procedures which are applicable in the fixed-covariate case (chapter 2) can also be used in case time-dependent covariates are present. For instance, the stability of the regression coefficients when the model is fitted in adjacent intervals of the time-axis can be investigated. It will also be possible to investigate whether regression-coefficients vary smoothly along time by incorporating terms representing functions of time itself in the regression model. Stratification on fixed covariates and estimating the cumulative baseline hazard functions, will allow assessment of the proportionality assumption by graphical means. The comparison of observed survival curves with estimated survival curves based on grouping of patients according to covariates does not apply here. These depend on past and future values of the time-dependent covariates which are unknown. By taking succesive short intervals, however, in which the covariate vector Z(t) is assumed to be constant during each interval, short term survival probabilities can be calculated [Hughes, 1992]. The probability of surviving time t+h conditional on being alive at time t can be approximated by

$$\exp\{e^{z(t)\beta}[-\hat{\Lambda}_0(t+h) + \hat{\Lambda}_0(t)]\} \qquad (6.3)$$

Here $\hat{\Lambda}_0(t)$ denotes the estimated cumulative baseline hazard function. For groupings of the covariate vector at time t, the short term observed survival probabilities can be compared with the estimated short term survival probabilities for each interval.
Also the plotting of residuals as defined by Barlow and Prentice (1988) may be useful. These residuals are computed for each subject separately for each variable and focus on the discrepancy between the covariate values at the failure time of the subject who dies and the covariate means of the corresponding risk set. These residuals can be plotted against the observation times. Although the plots are difficult to interpret, any trends or changes in variability are indicative of violations of the proportional hazards assumption [Altman and De Stavola, 1994].

6.5 Discussion and introduction to chapters 7-11

A key clinical question at diagnosis is that of prognosis. A means of updating prognosis on the basis of the latest observations on a patient would be extremely valuable in many situations. Also an assessment of factors related to the drop-out rate in clinical studies might be useful [Macera et al, 1988]. The proportional hazards model with updated covariates may be a useful tool for such analyses. The method allows the investigation of the history of covariates regarding their association with the instantaneous hazard rate. Its main use therefore lies in the description of the evolution of disease. Short-term predictions, that is predictions over periods of time in which the time-dependent factors do not vary greatly, are possible. Long-term predictions however are not possible. This would require an additional modelling of the time-path that the various covariates will follow [Mulder, 1993].
If the effect of treatment is of primary importance, as in the analysis of clinical trial data, it is extremely hazardous in general to allow for time dependent factors in the model. Because then there is a real danger that effects of treatment will be "adjusted away" by the inclusion of such factors. Similar difficulties regarding the interpretation will arise when other baseline covariates are evaluated regarding their prognostic

value while the statistical model also includes time-dependent covariates. For instance, a model relating various baseline characteristics in oncology to survival, at the same time allowing for metastases which may be found during follow-up of patients, will be hard to interpret. In exceptional cases, however, both the inclusion of treatment and time dependent variables in the same model may give useful results by illucidating the mode of action of the treatment. This is shown in one of the following chapters where data arising from a randomized clinical trial evaluating selective decontamination in acute pancreatitis are presented. In that study it was found that survival of patients was improved when using adjuvant selective decontamination. It was also found that in patients of the selective decontamination group less gram-negative infections of pancreatic necrosis occurred. Using time-dependent Cox-regression, allowing for both treatment and whether or not gram-negative infected necrosis had developed, it was found that a gram-negative infection of pancreatic necrosis was an ominous sign, while, taking account of this, no significant difference between the selective decontamination group and the control group remained regarding mortality.
In the following chapters various clinical studies are described in which the interest focussed on the assocation between factors as determined during follow-up and the occurrence of some endpoint. The above mentioned effect of a gram-negative infected pancreatic necrosis on mortality in acute pancreatitis is described in chapter 7.
It is often thought that transabdominal chorionic villus sampling during pregnancy may induce abortions. It goes without saying that such an effect can only be assessed be means of observational studies. In chapter 8 a study is described where the rate of spontaneous abortions is studied among women who were scheduled for such an examination. At the various gestational ages, the women who were still waiting for the procedure served as controls for the women who already had undergone the examination. Using Cox-regression the procedure related abortion risk was assessed.
In chapters 9, 10 and 11 (respective fields: cardiology, obstretics and oncology), clinical longitudinal studies are described in which monitoring measurements were evaluated regarding their ability to predict a defined endpoint.

References

Altman DG, BL de Stavola. Practical problems in fitting a proportional hazards model to data with updated measurements of the covariates. Statist Medic 1994; 13: 301-341.

Anderson JR, KC Cain, RD Gelber. Analysis of survival by tumor response. J Clin Oncol 1983, 1: 710-719.

Barlow WE, RL Prentice. Residuals for relative risk regression. Biometrika 1988; 75: 65-74.

Broekhoven P van, WCJ Hop, E Rasser, JC de Jongste, KF Kerrebijn. Comparison of FEV_1 and transcutaneous oxygen tension in the measurement of airway responsiveness to Metacholine. Pedr Pulm 1991; 11: 254-258.

Cox DR. Regression models and life tables (with discussion). J Roy Statist Soc 1972; B34: 187-220.

Crowley J. Asymptotic normality of a new nonparametric statistic for use in organ transplant studies. J

Am Stat Assoc 1974; 69: 1006-1011.
Crowley J, M Hu. Covariance analysis of heart transplant survival data. J Am Stat Assoc 1977; 72: 27-36.

Ettinger DS, S Lagakos. Phase III study of CCNU, cyclophosphamide, adriamycine, vincristine and VP-16 in small-cell carcinoma of the lung. Cancer 1982; 49: 1544-1554.

Gail MH. Does cardiac transplantation prolong life? A reassessment. Ann Int Medic 1972; 76: 815-817.

Gail MH. Evaluating serial cancer marker studies in patients at risk of recurrent disease. Biometrics 1981; 37: 67-78.

Hermans J, S Suciu, T Stijnen, P Aegerter, NC Gorin, A Gratwohl, M Hayat, P Stryckmans, R Zittoun, FE Zwaan. Treatment of acute myelogenous leukemia. An EBMT-EORTC restrospective analysis of chemotherapy versus allogeneic or autologous bone marrow transplantation. Eur J Cancer Oncol 1989; 25: 545-550.

Hughes MD, CL Raskino, SJ Pocock, MR Biagini, AK Burroughs. Prediction of short-term survival with an application in primary biliary cirrhosis. Statist Medic 1992; 11: 1731-1745.

Kalbfleisch JD, RL Prentice. The statistical analysis of failure time data. Wiley, New York, 1980.

Mantel N, DP Byar. Evaluation of response-time data involving transient states: an illustration using heart-transplant data. J Am Stat Assoc 1974; 69: 81-86.

Mauger EA, RA Wolfe, FK Port. Transient effects in the Cox proportional hazards regression model. Statist Medic 1995; 14: 1553-1565.

Macera CA, KL Jackson, C Farach, RR Pate. The use of proportional hazards regression in investigating dropout rates in a longitudinal study. J Clin Epid 1988; 41: 1175-1180.

McCullagh P, JA Nelder. Generalized linear models. Chapman and Hall, New York. 1989.

Mulder PGH. The simultaneous processes of ageing and mortality. Statistica Neerlandica 1993; 47: 253-268.

Oye RK, MF Shapiro. Reporting results from chemotherapy trials. Does response make a difference in patient survival. J Am Med Assoc 1984; 252: 2722-2725.

Prentice RL, JD Kalbfleisch, AV Peterson, N Flournoy, VT Farewell, NE Breslow. The analysis of failure times in the presence of competing risks. Biometrics 1978; 34: 541-554.

Redmond C, B Fisher, HS Wieand. The methodological dilemma in retrospectively correlating the amount of chemothearpy received in adjuvant therapy protocols with disease-free survival. Cancer Treatm Rep 1983; 67: 519-529.

Simon R, RW Makuch. A non-parametric graphical representation of the relationship between survival and the occurence of an event: application to responder versus non-responder bias. Statist Medic 1984; 3: 35-44.

Treurniet-Donker AD, WCJ Hop, S den Hoed-Sijtsema. Radiation treatment of stage III mammary carcinoma; a review of 129 patients. Int J Radiat Oncol Biol Phys 1980; 6: 1477-1482.

CHAPTER 7

DIFFERENTIAL PROGNOSIS OF GRAM-NEGATIVE VERSUS GRAM-POSITIVE INFECTED AND STERILE NECROSIS IN ACUTE PANCREATITIS

This chapter is submitted for publication by
E.J.T. Luiten, W.C.J. Hop, J.F. Lange and H.A. Bruining

Introduction

As infectious complications have become the leading cause of mortality and morbidity in acute necrotizing pancreatitis, patients in whom devitalized pancreatic and peripancreatic tissues remain sterile have to be distinguished from others in whom secondary infection of pancreatic necrosis develops (i.e. sterile vs. infected necrosis).
Sterile pancreatic necrosis, especially in the absence of systemic complications, has a favourable prognosis with a reported mortality rate of 0 - 11 %.[1-4] Mortality, which increases sharply with increasing Ranson or Imrie score on admission, is related to systemic complications resulting in multiple organ failure occurring most frequently during the first two weeks of illness.[5-10]
Infected necrosis on the other hand, occurring later during the course of the disease, often proves fatal and is generally agreed to represent an absolute indication for surgery in an effort to reduce mortality.[2,11-14]
Cultures of infected necrosis yield most frequently a polymicrobial flora with preponderance of gram-negative aerobic bacteria in 50 - 70 % suggesting an enteric origin.[2,5,12,15-19] Gram-positive aerobes (mainly *Enterococci* and *Staphylococci*) are isolated in only 5 - 20 %.[2,5,12,15-19]
Patients with severe acute pancreatitis in whom infected necrosis exists are usually dealt with as one group irrespective of the specific flora cultured. However due to a different intrinsic pathogenic potential, the prognosis of patients with infected pancreatic necrosis may differ according to the bacteria cultured. In accordance with results of bacteriological analyses of infected pancreatic necrosis two major groups of patients can be distinguished: gram-positive infected necrosis versus gram-negative infected necrosis. This distinction between patients with infected pancreatic necrosis has not been studied prospectively to date.
In a recent controlled clinical trial selective decontamination has been shown to effectively reduce mortality in patients with objective signs of severe acute pancreatitis.[5] Selective decontamination (SD) however does not prevent against gram-positive infection. Subsequent intestinal overgrowth with *Enterococcus* resulting in increasing gram-positive infections has been suggested to be a limitation of SD.[20]
The present prospective controlled clinical study was undertaken to evaluate for both treatment groups, i.e. SD group and C (= control) group, a possible difference concerning mortality between patients with either gram-positive infected necrosis or gram-negative pancreatic infection during the course of the disease, as compared to patients in whom pancreatic necrosis remained sterile.

Patients and methods

Between April 22, 1990 and April 19, 1993, 102 patients with objective signs of severe

acute pancreatitis were admitted to 16 participating hospitals. The diagnosis of acute pancreatitis had been established on the basis of clinical examination and elevated plasma levels of serum amylase (> 1000 IU per liter, normal range 0 - 300 IU/l (Phadebas)), or at diagnostic laparotomy (10 patients). All patients suffered from severe acute pancreatitis according to a multiple laboratory criteria score (Imrie score $\geq$ 3) and/or grade D or E disease severity (Balthazar grades) using contrast-enhanced computerized tomography.[7,21] Bacteriologically proven infected necrosis at the time of randomization was defined as an exclusion criterium. The patients were randomly assigned to receive standard treatment (C group : n = 52 pts.) or the same treatment plus selective decontamination (SD group : n = 50 pts.). A more elaborate outline has been reported previously.[5]

Microbiology:
The microbial flora of all patients was carefully monitored. Cultures from oropharynx, gastric content and rectum were taken on admission and repeated twice per week (surveillance cultures) until discharge. If fever ($\geq$ 39 ^{0}C) was present blood cultures were taken. An US- or CT-guided fine needle aspiration with subsequent culture was performed if there was clinical suspicion of infected pancreatic necrosis.[11] Cultures of pancreatic and peripancreatic devitalized tissues (i.e. necrosis) were obtained at every laparotomy. Drainage from the pancreatic bed was also cultured twice per week.
Samples from oropharynx, stomach and rectum were taken with a moistered sterile cotton-tipped swab and cultured semi-quantitatively. Specimens from the pancreatic necrosis were sent directly to the laboratory and cultured semi-quantitatively. Identification was done following routine microbiological procedures. Pancreatic necrosis, peripancreatic devitalized tissues and fluid collections were considered sterile in those patients pursuing a nonseptic course and in patients with negative cultures.

Surgery:
Surgery was performed if aspiration cultures demonstrated development of infected necrosis or if the condition of the patient was rapidly deteriorating towards a multiple organ failure resistant to exhaustive intensive treatment. Access towards the pancreas was obtained through a median or (preferably) transverse laparotomy. If repeated laparotomies were foreseen, a laparostomy, i.e. ventral open packing of the abdominal cavity, was created ensuring a rapid and easy access towards the upper abdominal cavity.[12] Removal of necrotic tissue was mainly performed by means of finger or clamp fraction, i.e. necrosectomy.

Statistical Analysis:
Percentages and continuous data were compared between groups by using Fisher's exact test and Mann-Whitney's test respectively. Cumulative percentages of patients developing gram-negative or gram-positive infected necrosis, taking account of the length of survival, were assessed by the actuarial Kaplan-Meier method and logrank-test. Cox regression was used to evaluate various factors simultaneously regarding mortality.[22] This method was also used to assess the relation between the occurrence of pancreatic infections (only gram-negative, only gram-positive or mixed gram-negative / gram-positive) and mortality.[23] P-values given are two-sided, and p = 0.05

was considered the limit of significance.

Results

Of 102 patients with objective signs of severe acute pancreatitis (Imrie score $\geq$ 3 and/or Balthazar CT score grade D or E) 50 patients were assigned to the SD group and 52 to the C group. The groups were well matched regarding Imrie score (both : mean 3.2) and Balthazar grade.[5]

Microbiology: twenty-nine of 102 patients (28 %) developed infected pancreatic necrosis. *Pseudomonas aeruginosa* (13 pts.), *Escherichia coli* (13 pts.), *Staphylococcus epidermidis* (21 pts.) and *Enterococci* (19 pts.) were most frequently isolated. Infected necrosis occurred in 9 of 50 patients (18 %) of the SD group in comparison to 20 of 52 patients (37 %) of the control group (p=0.03) due to a significant reduction of gram negative infected necrosis (SD : 4/50 patients (8 %); Control : 17/52 patients (33 %)). Figure 1 shows the increasing percentage of patients with time developing gram-negative pancreatic infection. The incidence of gram-positive infection of pancreatic necrosis did not significantly differ between treatment groups (SD : 9/50 patients (18 %); C : 16/52 patients (31 %)) (figure 1).

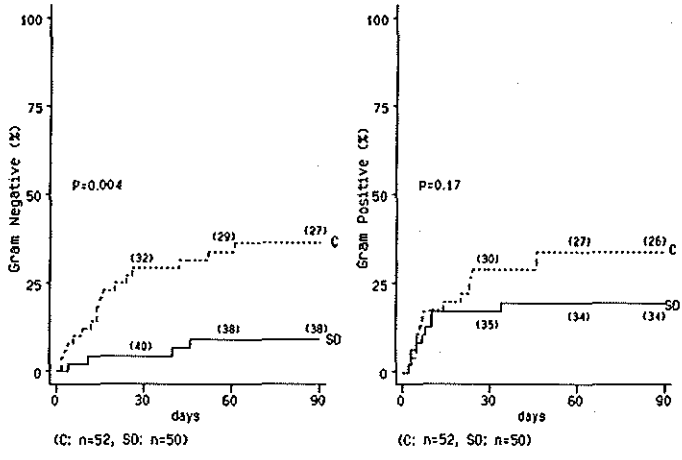


Figure 1. Cumulative (actuarial) percentage with time of patients with gram-negative infection (left panel) and with gram-positive infection of pancreatic necrosis (right panel) according to treatment group (Control (C) and SD). Numbers along curves indicate patients at risk.

The Imrie score at entry into the study appeared to correlate very strongly with the incidence of gram-negative pancreatic necrosis with time, especially in the control group ($p<0.001$) (figure 2). The Balthazar grade at entry into the study also correlated with the incidence of gram-negative pancreatic infection, although less pronounced (Grade C/D : 6/48 pts. (12 %) versus Grade E : 14/53 pts. (26 %))

(adjusted for treatment group : p=0.03).

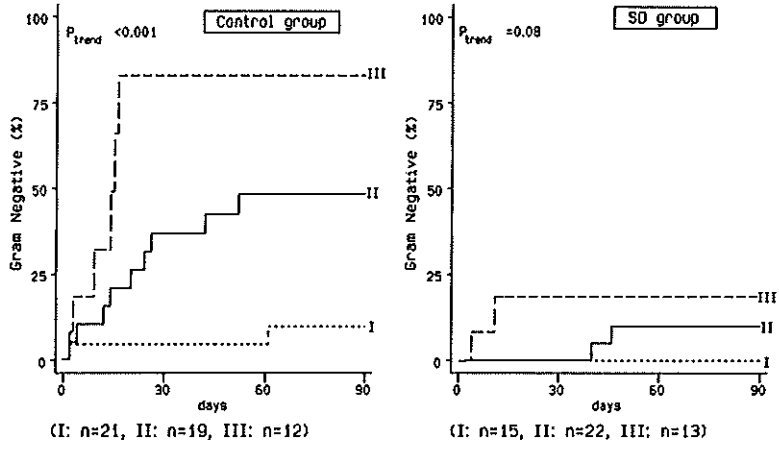


Figure 2. Cumulative (actuarial) percentage with time of patients with gram-negative infection of pancreatic necrosis in the C group (left panel) and SD group (right panel) according to Imrie score (I : 0-2 ; II : 3-4 ; III : > 4) at entry into the study.

Mortality: mortality is significantly reduced in patients treated with adjuvant SD.[5] In patients who survived (n = 73), the percentage of cases in whom infected necrosis had occurred was 19 % (14 pts.), which is significantly (p = 0.002) lower than the 52 % (15 pts.) of non-survivors (n = 29) (table 1).

Table 1: Bacteriological results of pancreatic necrosis during the course of the disease. Data given are numbers of patients.

	Survivors (n = 73)		Non-survivors (n = 29)	
Sterile	59	(81 %)	14	(48 %)
Only gram-positive	6	(8 %)	2	(7 %)
Only gram-negative	-	(-)	4	(14 %)
gram-positive/negative	8	(11 %)	9	(31 %)

To evaluate the impact on mortality of infected pancreatic necrosis developing during treatment, patients were classified each day according to whether pancreatic necrosis was still sterile, or whether only a gram-positive, only a gram-negative or a mixed gram-negative/gram-positive pancreatic infection had occurred. All patients started in the sterile condition in accordance with the entry criteria.

Table 2: Multivariate analysis of mortality in relation to the development of infection of pancreatic necrosis during treatment, baseline Imrie-score, baseline Balthazar grade and randomized treatment.

Data given between parentheses denote results when no allowance was made for the factor infections. All patients start in the sterile category in accordance with the entry criteria.

Factor	Deaths/ number*	Relative Death Rate		Significance†		95% confidence limits of relative death rate
Infected pancreatic necrosis						
Sterile	[14/5209]	1‡		-		-
Only G+	[2/615]	1.2		0.86		0.2, 5.5
Only G-	[4/103]	8.0§		0.001		2.4, 27
Mixed G+/G-	[9/508]	7.0§		0.002		2.1, 24.5
Imrie score						
0-2	[1/36]	1‡	(1)	-	(-)	-
3-4	[9/41]	7.4	(10.6)	0.06	(0.03)	0.9, 60.3
5-7	[19/25]	31.2	(56.8)	0.001	(0.001)	3.9, > 100
Balthazar Grade						
C/D	[9/48]	1‡	(1)	-	(-)	-
E	[20/53]	0.9	(1.3)	0.76	(0.38)	0.3, 2.3
Treatment						
Control	[18/52]	1‡	(1)	-	(-)	-
SD	[11/50]	0.8	(0.4)	0.56	(0.03)	0.3, 1.8

(G+: gram-positive; G-: gram-negative)

*: number denotes the number of patient-days (up to day 80, i.e. the day number of the last death) after the first occurrence of the infection specified, or the number of patients for Imrie score, Balthazar grade or treatment.

†: comparison with reference category

‡: reference category

§: not significantly different from each other, but both significantly greater in comparison with Only G+ category

{Missing data: Balthazar grade (n = 1)}

Using Cox-regression for both groups it emerged that, compared to patients with sterile necrosis, those who only acquired a gram-positive pancreatic infection had a 1.6 fold increased death rate ($p=0.52$) . Patients who developed only gram-negative pancreatic infection had a 14.4 fold increased death rate ($p<0.001$) compared to those with sterile necrosis. A similar increased mortality of 15.8 ($p<0.001$) was found in those with a mixed gram-negative/gram-positive infected necrosis.
Table 2 shows the results of the multivariate analysis of the relation between mortality and the type of pancreatic infection, taking into account the Imrie score, Balthazar grade and randomized treatment. This analysis demonstrates that development of a gram-negative infection of pancreatic necrosis during the course of the disease is both an important and ominous sign, while there was no significant increased mortality due to gram-positive pancreatic infection. Taking into account these infections, there is still an increased death rate of patients with a higher Imrie multifactorial initial assessment. No additional prognostic value was found for the Balthazar grade. The data between parentheses (table 2) also show that SD decreases mortality when analyzed without consideration of infectious status. There was no mortality difference between treatment groups for patients without gram-negative infected necrosis (i.e. sterile or only gram-positive) as demonstrated in figure 3 (left panel). After the occurrence of a gram-negative pancreatic infection mortality was high (13/21 pts. = 62 %). As shown in figure 3 (right panel) survival in these patients did not significantly differ between treatment groups. However a gram-negative pancreatic infection occurred in only 4 patients in the SD group.

Hospital stay: the average hospital stay in survivors with gram-negative necrosis ($n=8$ pts.) was 135 days (range 56-241 days), which is significantly higher than the mean value of 55 days (range 26-82 days) ($p=0.01$) as well as 30 days (range 10 - 71 days) ($p<0.001$) of survivors with only gram-positive ($n=6$ pts.) or sterile necrosis ($n=59$ pts.) respectively. Although smaller, the difference between hospital stay of survivors with only gram-positive infected and sterile necrosis is also significant ($p=0.004$). These results were similar in both treatment groups (SD or C group).

Discussion

This study demonstrates that mortality increases dramatically once gram-negative infection of pancreatic necrosis occurs in patients with severe acute pancreatitis. However if pancreatic necrosis becomes infected with only gram-positive aerobic bacteria mortality is not significantly increased and is comparable with patients in whom pancreatic necrosis remains sterile throughout the course of the disease, probably because *Staphylococcus epidermidis* or *Enterococci*, most frequently isolated in case of solitary gram-positive infected necrosis, are less pathogenic in these patients.
The overall incidence of secondary infection of pancreatic necrosis is 28 %. In the Control group 38 % infected necrosis occurred, which has also been described by others.[1,15,24] In patients treated with SD the overall incidence of infected necrosis (18 %) is significantly reduced due to a marked reduction of gram-negative infected necrosis of only 8 % in contrast to 33 % in the C group. The occurrence of a gram-negative infection of pancreatic necrosis is an ominous sign. Mortality in these patients increases significantly irrespective of coexistence of a gram-positive infection

as shown in this study. Mortality with regard to the bacteriological status of pancreatic necrosis is comparable for the SD group and the C group, i.e. once gram-negative infection of pancreatic necrosis has occurred mortality increases considerably in both treatment groups. However mortality in the treatment group is significantly reduced in patients treated with adjuvant SD (table 2) as has also been published previously.[5] Consequently, SD reduces mortality in patients with severe acute pancreatitis due to a significant reduction in the development of gram-negative infection of pancreatic necrosis. This is accomplished by reduction of gram-negative intestinal colonization leading to reduced gram-negative bacterial translocation into the pancreatic necrosis. However SD is not useful in patients in whom gram-negative pancreatic infection already exists before or in whom gram-negative infection develops during SD administration as is demonstrated in this study.

It has been suggested that overgrowth and translocation of gram-positive bacteria, i.e. *Enterococci* or *Staphylococci*, may be a draw back of SD.[20,25] Our results do not support this hypothesis. Neither intestinal overgrowth nor increased incidence of gram-positive infected necrosis has been found in our study. Mortality due to unexplained gram-positive sepsis with positive blood cultures in two patients and otherwise documented sterile necrosis at time of death, was equally divided among the SD and the C group. However these possible hazards demand strict indications and careful bacteriological surveillance as is the case for any kind of antibiotic regimen.

The development of gram-negative infection of devitalized tissues in and around the pancreas is, apart from the Imrie score, the most important parameter determining outcome. Gram-negative pancreatic infection can be minimized with adjuvant SD thereby reducing mortality in patients with severe acute pancreatitis.

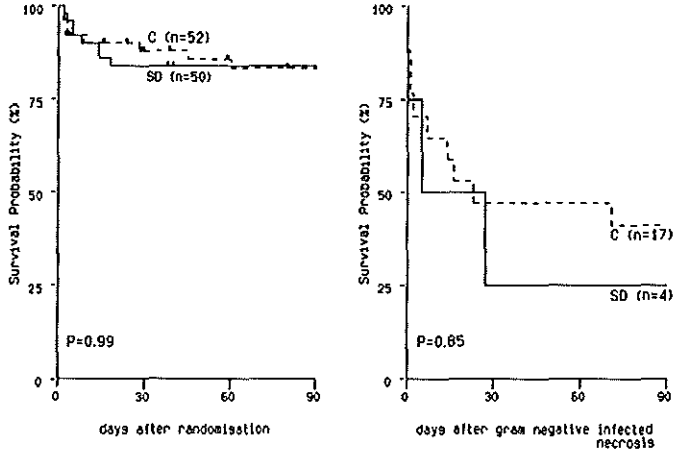


Figure 3. Left panel: Actuarial survival from entry into the study counting only deaths with sterile or only gram-positive infected pancreatic necrosis according to treatment group. Tick marks denote patients dying after a gram-negative infection of pancreatic necrosis. Right panel:Survival after the occurence of gram-negative infection of pancreatic necrosis according to treatment group.

References

1. Beger HG, Buchler M, Bittner R, Block S, Nevalainen T, Roscher R. Necrosectomy and postoperative local lavage in necrotizing pancreatitis. Br J Surg 1988; 75: 207-12.
2. Bradley EL 3d, Allen K. A prospective longitudinal study of observation versus surgical intervention in the management of necrotizing pancreatitis. Am J Surg 1991; 161: 19-24.
3. Rattner DW, Legemate DA, Meuller PR, Warshaw AL. Early surgical debridement of pancreatic necrosis is beneficial irrespective of infection. Am J Surg 1992; 162: 137-143.
4. Roscher R, Beger HG. Bacterial infection of pancreatic necrosis. In: Beger HG, Buchler M, eds. Acute pancreatitis. Springer-Verlag, New York. 1987: 314-320.
5. Luiten EJT, Hop WCJ, Lange JF, Bruining HA. Controlled clinical trial of selective decontamination for the treatment of severe acute pancreatitis. Ann Surg 1995; 222: 57-65.
6. Karimgani I, Porter KA, Langevin RE, Banks PA. Prognostic factors in sterile pancreatic necrosis. Gastroenterology; 103: 1636-1640.
7. Blamey SL, Imrie CW, O'Neill J, Gilmour WH, Carter DC. Prognostic factors in acute pancreatitis. Gut 1984; 25: 1340-1346.
8. Ranson JH. Etiological and prognostic factors in human acute pancreatitis: a review. Am J Gastroenterol 1982; 77: 633-638.
9. Steinberg W, Tenner S. Acute pancreatitis. New Eng J Med 1994; 330: 1198-1210.
10. Wilson C, Imrie CW. Systemic effects of acute pancreatitis In:Johnson CD, Imrie CW. eds. Pancreatic disease. Springer Verlag, London. 1991: 287-297.
11. Gerzof SG, Banks PA, Robbins AH et al. Early diagnosis of pancreatic infection by computed tomography-guided aspiration. Gastroenterology 1987; 93: 1315-1320.
12. Bradley EL. Management of infected necrosis by open drainage. Ann Surg 1987; 206: 542-550.
13. Ranson JH. The role of surgery in the management of acute pancreatitis. Ann Surg 1990; 211: 382-393.
14. Beger HG. Surgery in acute pancreatitis. Hepatogastroenterology 1991; 38: 92-96.
15. Beger HG, Bittner R, Block S, Buchler M. Bacterial contamination of pancreatic necrosis. A prospective clinical study. Gastroenterology 1986; 91: 433-438.
16. Widdison AL, Karanjia ND. Pancreatic infection complicating acute pancreatitis. Br J Surg 1993; 80: 148-154.
17. Pederzoli P, Bassi C, Vesentini S et al. Retroperitoneal and peritoneal drainage and lavage in the treatment of severe necrotizing pancreatitis. Surg Gynecol Obstet 1990; 170: 197-203.
18. Bittner R, Block S, Buchler M, Beger HG. Pancreatic abscess and infected pancreatic necrosis. Different local septic complications in acute pancreatitis. Dig Dis Sci 1987; 32: 1082-1087.
19. Bradley EL. Operative management of acute pancreatitis: ventral open packing. Hepatogastroenterology 1991; 38: 134-138.
20. Jackson RJ, Smith SD, Rowe MI. Selective bowel decontamination results in gram-positive translocation. J Surg Res 1990; 48: 444-447.
21. Balthazar EJ, Ranson JH, Naidich DP, Megibow AJ, Caccavale R, Cooper MM. Acute pancreatitis: prognostic value of CT. Radiology 1985; 156: 767-772.
22. Cox DR, Regression models and life tables. J Roy Stat Soc B 1972; 34: 187-220.
23. Hop WCJ, van Buuren HR. A method to evaluate changes in prognostic status during follow-up. Comp Biol Med 1989; 3: 181-188.
24. Smadja C, Bismuth H. Pancreatic debridement in acute necrotizing pancreatitis: an obsolete procedure? Br J Surg 1986; 73: 408-410.
25. Webb CH. Antibiotic resistance associated with selective decontamination of the digestive tract. J Hosp Inf 1992; 22: 1-5.

CHAPTER 8

SPONTANEOUS ABORTION RATES AND ADVANCED MATERNAL AGE: CONSEQUENCES FOR PRENATAL DIAGNOSIS

This chapter was published in The Lancet 1990; 336: 27-29
by T.E. Cohen-Overbeek, W.C.J. Hop, M. den Ouden, L. Pijpers, M.G.J. Jahoda and J.W. Wladimiroff.

Introduction

The primary indication for first-trimester prenatal chromosome analysis is advanced maternal age.[1,2] Jahoda et al[3] found that transcervical chorionic villus sampling (TC-CVS) at 9-11 weeks of gestation was associated with a significantly higher spontaneous abortion rate in women aged 36 and older (7·2%) than in women aged under 36 (2·6%). Whether this higher abortion rate is procedure related or due primarily to a higher baseline rate of fetal loss at advanced maternal age is uncertain. The objective of the present study was to determine maternal age related and procedure related fetal abortion rates in older women scheduled for transabdominal chorionic villus sampling (TA-CVS) at 12-14 weeks of gestation.

Patients and methods

384 pregnant women requesting prenatal diagnosis between January and October, 1988, were included in the study.
Admission criteria were: (i) maternal age 36 years or older at 20 weeks of gestation (ii) a viable singleton pregnancy; (iii) gestational age not beyond 76 days; (iv) no indication for prenatal diagnosis apart from advanced maternal age. TA-CVS was scheduled as an outpatient procedure between 12 and 14 weeks of gestation.[4] After reconfirmation of fetal viability by ultrasound, the skin was disinfected and a 20 gauge needle without stylet was introduced into the chorion frondosum under continuous ultrasound guidance. No local anaesthetic was used. The needle was attached to a 20 ml syringe filled with 5 ml saline. Chorionic tissue was obtained by moving the needle under continuous suction in a vertical and oblique direction. During each procedure the number of attempts and the amount of tissue collected were recorded. Non-attenders were contacted personally in order to establish the reason for failing their appointment.
In the event of a spontaneous abortion the gestational age at which this had occurred was documented. Continuing pregnancies were followed up by means of a questionnaire about short-term and long-term complications.

Statistics:
Cumulative percentages of women having a spontaneous abortion while being scheduled for TA-CVS were calculated with life-table methods[5]. The log-rank test was used to compare percentages between the different classes of gestational age and maternal age. In this analysis, women were considered removed from the study as

soon as TA-CVS had been completed.

To assess the effect of the TA-CVS procedure on the spontaneous abortion rate, patients themselves partially served as controls. The composition of the control group depended on the variable interval between intake and the TA-CVS procedure. For example, if 2 patients, matched for maternal and gestational age at intake, waited 15 and 20 days, respectively, then the interval between day 16 and day 20 for patient 2 would serve as the control period for patient 1, who by then had already undergone TA-CVS. Cox regression[6] was used to compare abortion rates between those still waiting and those who had already undergone TA-CVS at various points of time. In this analysis the day of gestation at intake was considered the day of entrance into the study. Relative (post versus pre TA-CVS) abortion rates, adjusted for maternal age, were calculated.[7] p values given are two-sided. The limit of statistical significance was set at $p<0\cdot05$.

Results

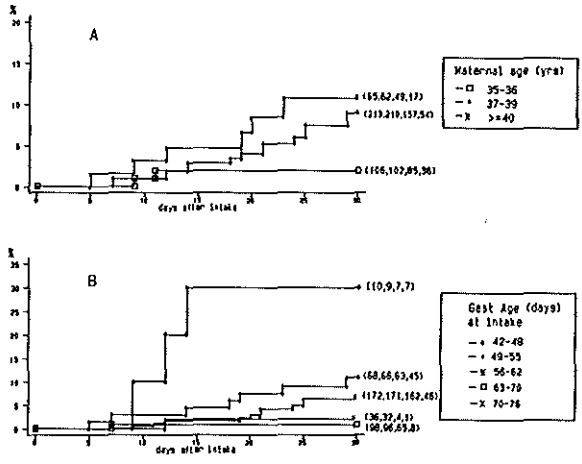
Maternal age of the 384 women ranged from 35 to 49 years (median 37 years) and maternal parity from 0 to 6 (median 1). Women were subdivided into five gestational age classes, 42-48 days (n=10), 49-55 days (n=68), 56-62 days (n=172), 63-70 days (n=98), and 71-76 days (n=36), and three maternal age classes, 35-36 years (n=106), 37-39 years (n=213), and 40 years or older (n=65). 5 women decided against prenatal diagnosis and 7 women did not undergo TA-CVS for technical reasons (retrochorionic haematoma, posterior placenta in a retroverted uterus, vaginal blood loss, HBAg positive). These women remained in the study up to the moment of withdrawal or cancellation.

There were 26 spontaneous abortions before TA-CVS. The percentage of women aborting while waiting for TA-CVS increased with advancing maternal age (figure, A). The percentage of women who aborted within 30 days after intake rose from 1·9% in the 35-36 years group to 10·9% in the group aged 40 and older (trend test:$p<0\cdot05$). Women entering the study between day 42 and 48 of gestation had a greater probability of aborting before TA-CVS than those entering beyond day 48 ($p<0\cdot05$; figure, B). The same observation was made when the comparison was adjusted for maternal age ($p<0\cdot01$). No correlation could be established between the incidence of spontaneous abortion and reproductive performance.

TA-CVS was carried out in 346 women, 6-54 days (median 26 days) after intake. Spontaneous abortion occurred in 6 women between 15 and 29 days after TA-CVS and in 2 women at 66 and 81 days. 11 pregnancies were terminated because of trisomy 21 (n=6), trisomy 18 (n=2), 47, XXY (n=2), or 47, XXX (n=1). The remaining 327 pregnancies progressed uneventfully up to 28 weeks. Subsequently, 1 woman suffered unexplained intrauterine death near term and 2 women were lost to follow-up.

The table shows the numbers of spontaneous and induced abortions before and after TA-CVS for each week of gestation in relation to maternal age. The women appeared to have a smaller risk of spontaneous abortion after than before TA-CVS, resulting in a maternal-age-adjusted relative abortion rate (post versus pre TA-CVS) of 0.10. This result, however, is severely biased because of the removal from the study of all chromosomally abnormal fetuses after TA-CVS. If all 19 abortions after TA-CVS are considered to have occurred spontaneously, the procedure led to a maternal-age-adjusted relative abortion rate of 1·66 (not statistically significant). Additional

adjustment for gestational age at intake resulted in a relative abortion rate of 1.56.



Cumulative percentages of women aborting spontaneously while awaiting TA-CVS according to (A) maternal age and (B) gestational age at intake. Numbers in parentheses denote number of women still being scheduled for TA-CVS on days 0, 10, 20, and 30 after intake.

Chorionic villus tissue was obtained after one sampling attempt in 307 women, including the 8 with spontaneous abortion after TA-CVS. In 1 woman the procedure was abandoned after one attempt, and a successful amniocentesis was carried out at 16 weeks. Two samplings were carried out in the remaining 38 women.

Discussion

The safety of first-trimester invasive procedures has been assessed against the background risk of spontaneous abortion. Incidences of 2·1% and 3·3% have been reported.[8,9] However, the percentage of women of advanced maternal age in these studies was always less than 9%.

Maternal age groups were differentiated in one study,[9] but the number of older women (over 40 years) was too small to allow any definite conclusions. The increase in chromosomal anomalies with advancing maternal age is more pronounced in populations undergoing CVS than in populations referred for amniocentesis.[10] This strongly suggests a maternal age dependent rise in fetal loss rate between 9 and 16 weeks of gestation. Moreover, the incidence of chromosomally normal abortions also rises with advancing maternal age, with a steep increase after 36 years.[11]

The present study confirms that there is a steady increase in early loss of a previously viable pregnancy with advancing maternal age. This must be taken into consideration when planning first-trimester chorionic villus sampling in older women.

Our study suggests that at advanced maternal age the risk of spontaneous abortion of

Table. Cumulated no of days patients were followed and no of abortions before and after TA-CVS for each week of gestation.

	Age 35-36 (n=106)				Age 37-39 (n=213)				Age >=40 (n=65)			
	Before TA-CVS		After TA-CVS		Before TA-CVS		After TA-CVS		Before TA-CVS		After TA-CVS	
Week	R	A	R	A	R	A	R	A	R	A	R	A
6	13	0	.	.	14	0	.	.	.	.	.	.
7	95	1	.	.	191	0	.	.	49	1	.	.
8	354	0	.	.	743	3	.	.	201	0	.	.
9	630	0	.	.	1186	1	.	.	358	1	.	.
10	705	1	6	0	1402	3	4	0	411	4	2	0
11	615	1	93	2*	1264	5	161	0	370	1	40	0
12	260	0	433	1*	500	3	882	2*	157	0	231	3*
13	59	0	621	1	125	0	1232	2*	46	0	332	0
14	8	0	659	1+1*	35	1	1300	1	27	0	351	0
15	.	.	658	0	25	0	1299	1	11	0	357	0
16	.	.	658	0	8	0	1299	1	.	.	357	0
17	.	.	658	0	.	.	1291	1	.	.	357	0
18	.	.	658	0	.	.	1288	0	.	.	357	0
19	.	.	658	0	.	.	1288	0	.	.	357	0
20	.	.	658	0	.	.	1284	1	.	.	357	0
21	.	.	658	0	.	.	1281	0	.	.	357	0
22	.	.	658	0	.	.	1281	1	.	.	357	0
23	.	.	658	0	.	.	1274	0	.	.	357	0
24	.	.	658	0	.	.	1274	0	.	.	357	0
Total abortions		3		2+4*		16		6+4*		7		3*

R=total number of days patients were monitored during a particular week of gestation.
A=no of abortions. *induced abortions after TA-CVS

a previously viable fetus in the first 12 weeks of gestation is greater than the chance of live birth of an infant with a chromosomal anomaly. The decline in abortion rate is more pronounced between 6 and 7 weeks of gestation than later. The increased abortion risk for women entering the study in week 6 could be explained by the apparently high initial abortion risk for these women during weeks 7 and 8. At that time the abortion rate was 0·024(3/124) abortions per day, whereas during the same period the abortion rate for women entering the study during week 7-8 was 0·001 (2/1509) abortions per day. This reduction in abortion rate is probably influenced by the ultrasound examination carried out at the intake visit. The exclusion, at that stage, of all non-viable pregnancies would explain the low abortion rate in patients entering during week 7-8.

A relation between gestational age at intake and abortion rate has also been observed in other studies in which early fetal viability was documented by ultrasound.[12,13] The only maternal age dependent data on fetal loss rates at given gestational periods presented in life-table style[14] were based on a cross-sectional study design and are therefore difficulty apply to a cohort under observation at a given stage of pregnancy. Moreover, in that study fetal viability was not verified by ultrasound. From the present study it has also become evident that a considerable number of pregnancy losses before TA-CVS occurred at 10-12 weeks of gestation.
This coincides with the period surrounding the TC-CVS procedure and explains the relatively high spontaneous fetal loss rate after TC-CVS in the advanced maternal age group.[3]
In contrast to Regan,[15] we were unable to establish any relation between spontaneous abortion rate and past reproductive performance. This may be because non-viable pregnancies were disregarded in the present study; 50% of the spontaneous abortions in Regan's study would have fallen into this category.
We found no evidence for any adverse effect of TA-CVS on the abortion rate. The abortion rate after TA-CVS was not significantly higher than that before the procedure, even if it is assumed that all abortions would have occurred spontaneously. However, when all diagnosed chromosomal anomalies were excluded, the spontaneous abortion rate before TA-CVS was considerably higher than that observed after the procedure.
We suggest that these findings justify late first-trimester chronic villus sampling in women of advanced maternal age because the spontaneous abortion rate and the procedure related abortion risk do not exceed the risk of fetal chromosomal abnormality. At 12 weeks of gestation termination is still feasible as an outpatient procedure. This approach will also be more cost effective, since a number of affected fetuses diagnosed early in the first trimester most likely would have aborted spontaneously by this time.

References

1. Brambati B, Lanzani A, Oldrini A. Transabdominal chronic villus sampling. Clinical experience of 1159 cases. Prenatal Diagnosis 1988; 8: 609-617.
2. Sachs ES, Jahoda MGJ, Kleyer WJ, Pijpers L, Galjaard H. Impact of first trimester chromosome, DNA and metabolic studies on pregnancies at high genetic risk: experiences with 1000 cases. Am J Med Genet 1988; 29:293-303.
3. Jahoda MGJ, Pijpers L, Vosters RPL, Wladimiroff JW, Reuss A, Sachs ES. Role of maternal age in assessment of risk of abortion after prenatal diagnosis during first trimester. Br Med J 1987; 295: 1237.
4. Pijpers L, Johoda MGJ, Reuss A, Wladimiroff JW, Sachs ES. Transabdominal chorionic villus biopsy in second and third trimesters of pregnancy to determine fetal karyotype. Br Med J 1988; 297: 822-823.
5. Peto R, Pike MC, Armitage P, et al. Design and analysis of randomized clinical trials requiring prolonged observation of each patient (part II: analysis and examples). Br J Cancer 1977; 33: 1-39.
6. Cox DR. Regression models and life-tables. J R Statist Soc 1972; B34: 187-220.
7. Hop WCJ, van Buuren HR. A Method to evaluate changes in prognostic status during follow-up. Comp Biol Med 1989; 19: 181-188.
8. Christiaens GCML, Stoutenbeek PH. Spontaneous abortion in proven intact pregnancies. Lancet

1984; ii: 571-572.
9. Gilmore DH, McNay MB. Spontaneous fetal loss rate in early pregnancy. Lancet 1985; i: 107.
10. Hook EB, Cross PhK, Jackson L, Pergament E, Brambari B. Maternal age-specific rates of 47, +21 and other cytogenetic abnormalities diagnosed in the first trimeser of pregnancy in chorionic villus biopsy specimens: Comparison with rates expected from observations at amniocentesis. Am J Hum Genet 1988; 42: 797-807.
11. Stein ZA. A women's age: childbearing and child rearing. Am J Epidemiol 1985; 121: 327-340.
12. Liu DTY, Jeavons B, Preston C, Pearson D. A prospective study of spontaneous miscarriage in ultrasonically normal pregnancies and relevance to chorion villus sampling. Prenatal Diagnosis 1987; 7: 223-227.
13. Steer C, Campbell S, Davies M, Mason B, Collins W. Spontaneous abortion rates after natural and assisted conception. Br Med J 1989; 299: 1317-1318.
14. Gustavii B. Chorionic biopsy and miscarriage in first trimester. Lancet 1984; i: 562.
15. Regan L, Braude PR, Trembath PL. Influence of past reproductive performance on risk of spontaneous abortion. Br Med J 1989; 299: 541-545.

CHAPTER 9

CYTOIMMUNOLOGICAL MONITORING OF HEART TRANSPLANT RECIPIENTS

This chapter was published in Clinical Transplantation 1990; 4: 297-300 by
N.H.P.M. Jutte, W.C.J. Hop, R. Daane, C.E. Essed, W. Weimar, M.L. Simoons and E. Bos

Introduction

Monitoring of rejection after clinical heart transplantation is generally performed by judging histological sections of endomyocardial biopsies (EMB) which are taken via the jugular vein[1]. Reports have appeared describing a method by which rejection crises could be observed in peripheral blood. The advantages of a non-invasive method for the patient are clear. In addition, this method, called cytoimmunological monitoring (CIM), was reported to be very sensitive so that rejection in the CIM could be observed 2-3 days before signs of rejection were observed in the EMBs[2]. The positive results obtained[2,3] prompted us to investigate the value of CIM for the population of heart recipients in Rotterdam, although negative results of CIM have also been published[4].
In the present study we investigated the prognostic value of CIM with respect to the first rejection episode during the first 3 months after transplantation. A randomized clinical trial evaluating prophylactic immunosuppression with either OKT3 or cyclosporin was started during the period of investigation. CIM results within each of these groups were also evaluated.

Patients and methods

Seventy patients transplanted between January 1985 and December 1988 were included in the study.
The prophylactic immunosuppressive therapy consisted of either cyclosporin or the monoclonal antibody OKT3 (Orthoclone 5 mg/d i.v. for 7 d) together with 50 mg azathioprine per d, both with 60 mg prednisone per d, decreasing to 10 mg at 7 weeks posttransplantation, starting immediately after transplantation, followed by cyclosporin and low-dose steroids. Many of the patients (45) participated in randomized clinical trial to evaluate both regimens. EMBs were taken once a week until 6 wk postoperatively, once every 2 wk until 3 months after transplantation, once a month until 6 months posttransplantation, once every 6 wk until 9 months posttransplantation, once every 8 wk until 12 months and once every 4 months thereafter. CIMs were made concurrent with EMB during 3 months after transplantation.
At the day of EMB, peripheral blood was separated over a Ficoll-isopaque gradient to obtain a concentrate of mononuclear cells. The cells were washed, a cytospin slide was made and stained with May-Grünwald Giemsa according to standard methods. The cells were counted, differentiating between lymphoblasts, activated lymphocytes, normal lymphocytes, large granular lymphocytes, juvenile granulocytes, juvenile neutrophil leucocytes ("rod-like") and plasma cells based on morphological appearance. Three-hundred cells of the above cell types were counted on at least 3 different sites in the cytospot. The number of lymphoblasts and activated lymphocytes was

calculated as percentage of the total amount of lymphocytes. The slides were counted without knowledge of the histological score of EMB.
A rejection episode was defined as moderate rejection according tot the criteria of Billingham[5], i.e. when infiltrates and myocyte necrosis were observed. Only the first rejection episode of each patient was analyzed in this study as subsequent rejection therapy might influence the results thereafter. In the present group of patients a severe rejection episode, described as myocyte necrosis and interstitial bleeding, was not observed. All patients were considered evaluable for CIM. Excluded were patients who died within 2 wk posttransplantation or who were transplanted elsewhere although being monitored at our center.
Statistical methods:
The relation between immunological parameters and the risk of first rejection as demonstrated by EMB was investigated by Cox-regression[6] with CIM data as time-dependent variables. p-values given in these analyses are derived from likelihood-ratio tests. Parameters within patients were also compared using Wilcoxon's test. Five percent (two-sided) was considered the limit of statistical significance.

Results

A total of 371 CIMs were made from 70 patients concurrently with EMB within a period of 3 months after transplantation or, if a rejection occurred within this period, up to the date of rejection. During this period 35 patients showed an episode of rejection concurrent with investigation by CIM. An additional 5 patients rejected within 90 d posttransplantation but no CIM was available at the day of rejection. CIM data of these 5 patients were included in the evaluation up to the latest measurement at the day of EMB preceding the day of rejection. Fourteen of the 35 CIM-monitored rejections occurred within 1 month after operation. Nine patients had a first rejection episode after 90 d post-transplant. Twenty-one patients did not have a rejection episode within the first 3 months nor in the follow-up period of at least 6 months thereafter, although 1 of these patients died of causes not related to rejection at 5 months post-transplant.
The median values of all CIMs counted concurrent with EMBs indicating no rejection were 0.5% (range: 0-13.4) lymphoblasts, 2.7% (range: 0-51.8) activated lymphocytes and 3.2% (range: 0-55.2) for the sum of both counts (n=336). Similar values were obtained for CIMs counted concurrently with EMBs indicating rejection within 90 d. For this group, median values were 0.5% (range: 0-2.9) lymphoblasts, 2.4% (range: 0-17.2) activated lymphocytes and 3.5% (range: 0-17.2) for both counts summed (n=35). Figure 1 shows, for each posttransplantation day at which rejection was observed, the counts of lymphoblasts and activated lymphocytes for patients with rejection in comparison to counts for patients in whom CIM was made at exactly the same post-transplant day but in whom no rejection was found. Cox-regression with time-dependent variables of the data given in this figure did not show an increased risk of rejection when larger percentages of lymphoblasts or activated lymphocytes were observed, either considered separately or in combination.

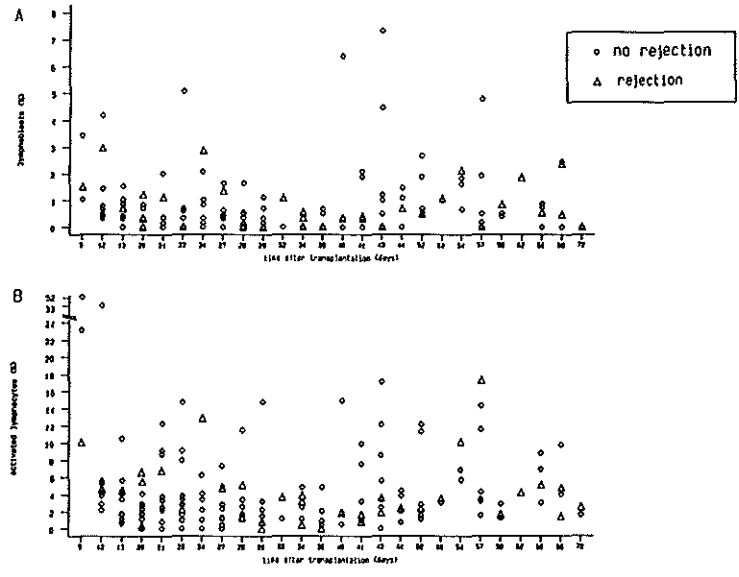


Figure 1. Percentages of lymphoblasts (A) (upper panel) and activated lymphocytes (B) (lower panel) in peripheral blood of heart transplant recipients according to time after transplantation. Represented are values of all individual patients who were measured on the day posttransplant at which 1 or more patients experienced rejection.

The estimate of the relative risk of rejection associated with an increase of the percentage of lymphoblasts with 1% was 0.88 which, however, is far from statistical significance ($p=0.6$) as compared to 1 (i.e. no relation). The relative risk associated with an increase of the percentage of activated lymphocytes with 1% was estimated to be 0.96 (compared to 1; $p=0.4$). Both estimates of the relative rejection risk did not significantly differ between the first period of 1 month after transplantation and the period thereafter.

Any correlation of CIM and rejection was also not apparent when analyses were performed separately for each immunosuppresive regimen used. All findings applied equally when the number of control-observations for each day at which rejection was observed was increased by using interpolated values of the counted immunological parameters of patients without rejection (data not shown in Figure 1).

When, instead of comparing CIM values between patients or groups of slides, changes in values within individual patients are considered, a different evaluation can be made. Figure 2 shows counts of lymphoblasts and activated lymphocytes in individual patients at the time of rejection in comparison to the counts at the latest CIM preceding the rejection. The median time interval between both CIMs was 8 d. Both parameters, and their sum, were not significantly increased at rejection as compared

to the previous measurement. In only 1 patient was a large increase in lymphoblasts observed at rejection. In another patient the percentage of activated lymphocytes had increased greatly (Figure 2). When time posttransplantation was included in evaluating these data by applying Cox-regression to study the pattern of increases/decreases within patients with time, no relation with rejection was apparent. This was also the case for each immunosuppressive regimen separately.

Discussion

The present results show that in our hands CIM is not useful in replacing EMB for diagnosis of rejection. The results are in contrast with previous claims that CIM is a help in detecting rejection at an early stage[2,7] and that the amount of lymphoblasts is most significant in this respect[8]. Differences in immunosuppression regimens, incidence or severity of rejection may cause the different results. In the present study no prognostic value of CIM could be found for two different immunosuppression regimens.

The first rejection episode appears to be most apt for evaluation since no interference from anti-rejection therapies can occur in the comparison of the EMB and the CIM. In addition, patients who reject more than once should not weigh more in the analysis. Therefore, in the present study only the first rejection in our group of patients were analyzed. Because most first rejections take place within 3 months after transplantation, we analyzed in the present study the CIMs obtained within this period of 90 d. As in other studies[7,8] activation of the immune system by bacterial or viral infections was not excluded.

Different CIM studies appear not be directly comparable since either the concentrations of activated cells were given[2,3,7] or percentages of activated lymphocytes were used[8]. It was demonstrated that use of concentration or percentage did show similar results with respect to detection of rejection[4,9,10]. In the present study analysis was based on the amount of activated cells as percentage of the amount of lymphocytes as has been done previously by the München group[8].

The percentages of activated cells in our studies are low as compared to the findings of others. The level of 6% lymphoblasts (as percentage of the amount of lymphocytes) which was reported as a limit for discrimination between non-rejection and rejection[8] was hardly ever reached in our material. In the present group of patients a severe rejection was never observed as judged by EMB, either in the first rejection or in later rejection episodes. In groups of patients studied by others, the rejecting group appeared to include both moderate and severe rejection[7,8]. Thus the difference between rejecting and non-rejecting patients may be larger in other studies since more severe rejections appeared to be observed. As shown in the present results, higher amounts of activated cells were observed incidentally.

It has been reported that the greatest sensitivity of the method is reached during the 1st month following operation[7]. However, our present data do not support these findings.

In conclusion, monitoring of rejection of heart transplant recipients by morphological

analysis of peripheral blood by CIM was not useful for the group of patients transplanted in Rotterdam.

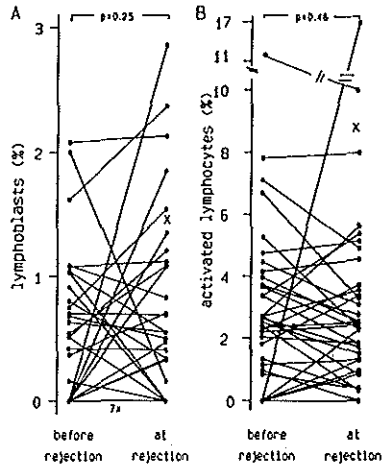


Figure 2. Percentages of lymphoblasts (A) or activated lymphocytes (B) in individual patients at the day of rejection as compared to the last preceding measurement (before rejection). Pairs of values for each patient are connected, except for one patient (x) who showed rejection in the first biopsy after transplantation. One line (7x) represents seven patients.

References

1. Caves PK, Stinson EB, Graham AF, Billingham ME, Ghrehl TM. Percutaneous transvenous endomyocardial biopsy. Am Med J 1973; 225: 228.
2. Hammer C, Reichenspurner H, Ertel W, et al. Cytological and immunologic monitoring of cyclosporine-treated human heart recipients. Heart Transplant 1984; 3: 228.
3. Reichenspurner H, Kemkes BM, Osterholzer G, et al. Special control of infection and rejection episodes after four years of cardiac transplantation at the University of Munich. Tex Heart Inst J 1986; 13: 5.
4. Hanson CA, Bolling SF, Stoolman LM, et al. Cytoimmunological monitoring and heart transplantation. J Heart Transplant 1988; 7: 424.
5. Billingham ME. Diagnosis of rejection of endomyocardial biopsy. J Heart Transplant 1982; 1: 25.
6. Cox DR. Regression models and life tables. J R Stat Soc 1972; B34: 187.
7. Fieguth HG, Haverich A, Schäfers HJ, et al. Cytoimmunologic Monitoring in early and late cardiac rejection. J Heart Transplant 1988; 7: 95.
8. Klanke D, Hammer C, Dirschedl P, et al. Sensitivity and specificity of cyto-immunological monitoring in correlation with endomyocardial biopsies in heart transplant patients. Transplant Proc 1987; 19: 3781.
9. Wijngaard PLJ, Heyn A, van der Meulen A, et al. Cytoimmunologic monitoring following heart transplantation. Cardiology 1987; 74: 415.
10. Wijngaard PLJ, van der Meulen A, Schuurman HJ, et al. Cytoimmunologic monitoring for the diagnosis of acute rejection after heart transplantation. Transplant Proc 1989; 21: 2521.

CHAPTER 10

THE PREDICTIVE VALUE OF DOPPLER FLOW VELOCITY WAVEFORMS IN THE DEVELOPEMENT OF ABNORMAL FETAL HEART RATE TRACES IN INTRAUTERINE GROWTH RETARDATION; A LONGITUDINAL STUDY

This chapter was published in Early Human Developement 1993; 32: 151-159 by I.A.L. Groenenberg, W.C.J. Hop, J.W. Bogers, J.G. Santema and J.W. Wladimiroff

Introduction

Less than 40% of pregnancies complicated by intrauterine growth retardation (IUGR) are recognized by maternal signs or symptoms which prompt the clinician to suspect IUGR[14]. Serial biometric measurements by ultrasound have improved the pick-up rate of IUGR. However, not all those small fetuses are necessarily at risk of becoming hypoxaemic. Doppler flow velocity waveform studies in the fetus have provided characteristic changes in various fetal vessels relative to IUGR [7,8,17,19]. Comparative studies between umbilical artery flow velocity waveform recordings and standard antepartum fetal heart rate testing suggest the former technique to be a valuable adjunct in the antepartum fetal surveillance in high-risk pregnancies[13]. Despite its limitations, most centres still consider fetal heart rate monitoring as the method of choice in the assessment of fetal compromise. No studies are known in which the time-relationship between these two methods of fetal surveillance was determined in a longitudinal blinded study design. The objective of the present study was to assess the predictive value of serial cardiac and extra-cardiac blood flow velocity waveform recording in the prediction of an abnormal fetal heart rate (FHR) trace in IUGR.

Materials and methods

Forty-two out of 43 women with a singleton pregnancy admitted for IUGR consented to participate in the study. The study protocol was approved by the Hospital Ethics Review Board. IUGR was defined as a fundal growth delay of more than 2 weeks and a deviation of the sonographic fetal upper abdominal circumference to beneath the 10th centile of the reference curve. Pregnancy duration was determined from the last menstrual period and confirmed by sonographic measurement of the fetal biparietal diameter at 14-20 weeks gestation. In 33 women pregnancy was further complicated by pregnancy-induced hypertension. All women were entered into the study within one to 2 days of hospital admission. Gestational age at the entry to the study varied between 26 and 35 weeks of gestation (median 31 weeks). Clinical management consisted of bedrest and treatment of hypertension. The appropriate time of delivery was determined by daily monitoring of fetal heart rate and maternal condition thereby taking into account gestational age.
Doppler flow velocity waveform recordings were made by an independent observer (IALG) on the day of entry into the study (day 1) and subsequently at 2-3 day intervals until delivery. Maximum flow velocity waveforms were recorded at both

cardiac and peripheral level, using a combined mechanical sector and pulsed Doppler system (Diasonics CV 400). Flow velocity waveforms from the fetal ascending aorta (AO) were obtained from the five chamber view[8]. Fetal pulmonary artery (PA) flow velocity waveforms were recorded from the conventional echocardiographic short axis view[8]. Doppler sample volumes were placed in the great vessels immediately distal to the semilunar valves. The angle between the Doppler cursor and the assumed direction of flow was always 10 degrees or less. Sample volume length was between 0.2 and 0.4 cm. Peak systolic velocity (PSV, cm/s) and time-averaged velocity (AV, cm/s) were calculated. Flow velocity analysis consisted of tracing the outer border of the densest part of the Doppler spectrum envelope. For both parameters a good reproducibility was previously established[9]. Peripheral arterial Doppler studies were focused on the fetal umbilical artery (UA) and the fetal internal carotid artery (ICA). For both vessels, the pulsatility index (PI) was calculated through dividing the difference between the peak systolic velocity and end-diastolic velocity by the time-averaged velocity.

All flow velocity waveforms were recorded on hard copies for analysis after delivery by a microcomputer (Olivetti M24) linked to a graphics tablet. An average of four consecutive flow velocity waveforms of good quality was used to establish each value.

Fetal cardiotocograms were recorded daily at a median duration of 35 minutes (range 25-60 minutes). In case of a flat FHR trace, recordings were made up to 60 minutes to exclude the "non-reactive" FHR pattern inherent to fetal rest-activity cycles. After delivery, all FHR traces were classified by two experienced independent observers (JWB and JS), who had no knowledge of the Doppler outcomes, according to the Fischer score[6]. This score is determined by five variables each being classified as good (2 points), moderate (1 point) or poor (0 points). These variables are: (i) heart rate, (ii) heart rate band width, (iii) number of zero-crossings per minute of baseline fetal heart rate, (iv) heart rate accelerations, and (v) heart rate decelerations. When the scoring difference between the two observers was 3 points or more, the FHR trace was evaluated by a third independent observer, whereby the median value was determined (13 recordings). In case of cyclic activity, assessments were only made during the active sleep state[15]. An FHR trace was considered abnormal when the Fischer score totaled 6 or less. Considering that Doppler flow measurements cannot predict an abnormal FHR trace when the latter is related to an obstetric intervention (n=3), in the analysis, the Doppler data of these three patients were only evaluated up to the day of cervical priming.

All Doppler parameters were expressed as standard deviation scores. This score is obtained by taking the difference between the observed value and the mean reference value according to the gestational age and dividing the result by the standard deviation of the reference values[10,20]. Logistic regression[3] was used to investigate the relationship between Doppler parameters and the presence of an abnormal FHR trace on the day of entry into the study (day 1). For patients who did not show an abnormal FHR trace on day 1, Cox-regression with time-dependent variables was used to investigate the relation between the Doppler parameters and the first occurrence of an abnormal FHR trace during follow-up[4]. Gestational age served as the time-axis in this analysis. Cumulative percentages of an abnormal FHR trace during follow-up were determined by actuarial methods[11].

Results

The number of Doppler sessions per subject varied between 1 and 20 (median 8), totaling 247 Doppler sessions and 988 Doppler measurements. The success rate in obtaining technically acceptable Doppler flow velocity waveforms in the umbilical artery, internal carotid artery, pulmonary artery and ascending aorta was 100%, 96%, 94% and 87%, respectively. One woman dropped out two weeks after hospital admission, due to referral to another hospital. Doppler and heart rate data obtained during these weeks were included in the analysis. Fetal birth weight varied between p<2.3 and p25 (median p5) according to Kloosterman's Tables corrected for maternal parity and fetal sex[12]. Caesarean section was performed in 36 women, vaginal delivery took place in the remaining 6 women (including one intrauterine death). Umbilical artery pH after delivery ranged between 6.82 and 7.33 (median 7.20). Umbilical artery BE ranged between -0.9 and -24.2 (median -7.0).
Fifteen out of 42 patients displayed an abnormal FHR trace on day 1 (day of entry into the study). The presence of an abnormal FHR trace on day 1 correlated significantly with gestational age: patients with an abnormal FHR trace were of a significantly (Mann-Whitney's test: p=0.03) lower gestational age than those with a normal FHR trace (mean 29.8 weeks, SD 1.8 versus mean 31.3 weeks, SD 2.5). Table I presents data on the various Doppler parameters, expressed as standard deviation scores according to whether or not abnormal FHR traces were present on day 1. When taking into account gestational age, PIUA and PSVAO scores separately showed a significant correlation with the occurrence of an abnormal FHR trace on day 1. Simultaneous evaluation of these flow parameters, however, demonstrated that only the PIUA score remained significantly correlated. The risk for an abnormal FHR trace was increased by 20% (P <0.01) for an increase in PIUA of one standard deviation score.

Longitudinal data from the remaining 27 patients (788 Doppler measurements) who had a normal FHR trace on day 1, were used to establish the predictive value of Doppler recordings for the development of an abnormal FHR trace during the course of pregnancy. Fourteen patients developed abnormal FHR traces, 13 patients did not. All 14 patients developing abnormal FHR traces were delivered by caesarean section, whereby the time delay between the occurrence of an abnormal FHR trace and delivery ranged between 0 and 15 days (in 71% of cases this interval was less than or equal to 2 days). Five out of 13 patients who did not develop abnormal FHR traces were delivered vaginally. In the remaining eight women caesarean section was performed because of partial abruptio placentae (n=2), severe pre-eclampsia (n=4) or gestational age of 36 weeks or more in the presence of non-optimal FHR traces (n=2). Cox-regression revealed that the PIUA, PIICA, PSVAO and AVAO scores were significantly correlated with the incidence of an abnormal FHR trace. However, when each parameter was evaluated separately from simultaneous assessment of these variables, it appeared that the correlation only remained signifcant for the PIUA and PIICA scores. Relative rates of an abnormal FHR trace for both parameters are given in Table II.

Table I. Mean scores of Doppler parameters in the presence or absence of an abnormal FHR trace on the day of entry into study.

	abnormal FHR trace present			absent			
Doppler parameter	mean	sd	n	mean	sd	n	p*
PIUA	12.0	11.0	15	3.7	4.0	27	.001
PIICA	-2.9	1.1	15	-2.4	1.5	26	.09
PSVAO	-1.6	0.9	14	-0.9	1.4	22	.04
AVAO	-1.4	1.0	14	-0.8	1.1	22	.08
PSVPA	-2.8	1.8	14	-2.2	1.5	23	.16
AVPA	-2.2	1.7	14	-1.8	1.3	23	.21

(*) significance, adjusted for gestational age (logistic regression). Total number of patients not always 42 due to different success rates for various Doppler measurements. PIUA, PIICA = pulsatility index in the umbilical artery and internal carotid artery. PSVAO, AVAO =peak systolic and time-averaged velocity in ascending aorta. PSVPA, AVPA = peak systolic and time-averaged velocity in pulmonary artery.

Table II. Multivariate analysis of PIUA and PIICA scores with respect to the incidence of an abnormal FHR trace.

Variable	relative rate of developing an abnormal FHR trace	P-value	confidence limits (90%)
PIUA	1.5[a]	.01	1.2-1.9
PIICA	2.6[b]	.05	1.3-5.4

As compared with fetuses who dispay values 1 standard deviation score lower for PIUA (a), or 1 standard deviation score higher for PIICA (b).

The multivariate assessment of the PIUA and PIICA scores resulted in a prognostic index for the development of an abnormal FHR trace during follow-up: 0.43 x PIUA score - 0.96 x PIICA score. Figure 1 depicts the values of this index for the 27 patients during follow-up. Data points for the 14 patients developing an abnormal FHR trace are interconnected, with the last data point representing the Doppler measurement obtained within 24 hours prior to the occurrence of an abnormal FHR trace. As can be seen, the prognostic index values of patients developing an abnormal

FHR trace tend to be higher than those who did not. The prognostic index appears to increase with advancing gestational age. By fitting a least-squares regression line (A) through the data points of patients who did not develop an abnormal FHR trace, the rising trend of the prognostic index was characterized. The prognostic value of the index was evaluated for two arbitrary chosen cut-off levels, drawn 1 SD (B) and 2 SD (C) above the regression line, respectively (Fig 1).
The actuarial cumulative percentage of patients who developed an abnormal FHR trace was determined I) according to the number of days past the day on which prognostic index values from patients were situated for the first time above the chosen cut-off levels, B (+1SD) and C (+2SD) respectively, and II) according to the number of days past the day of entry into the study prognostic index values remained below the chosen cut-off levels. Figure 2 displays the actuarial percentages for these two cut-off levels. For the +1SD cut-off level, the cumulative percentage of patients who had developed an abnormal FHR trace amounted to 47% within 1 week and to 77% within 4 weeks after the day the prognostic index value was first found to be situated above this level (curve II, upper panel, based on n=18 patients). As for the +2SD cut-off level, within 3 weeks after the first measurement was situated above this level the cumulative percentage of patients developing an abnormal FHR trace rose to 100% (curve II, lower panel).

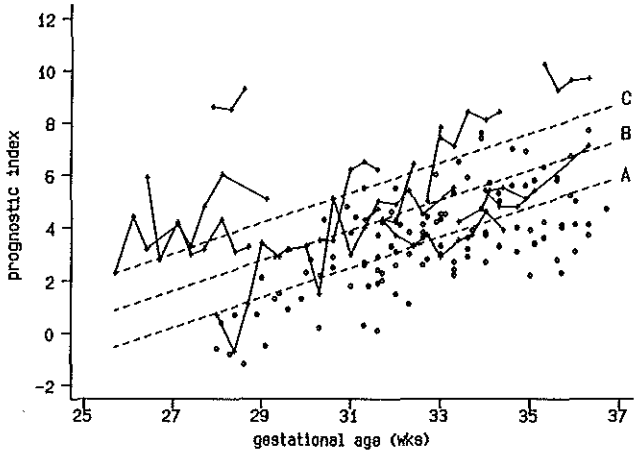


Figure 1. Prognostic index values against gestational age. Data points of patients (n=14) developing an abnormal FHR trace are interconnected. Data points of patients (n=13) who never developed an abnormal FHR trace are given as individual data points. (A) Least-squares regression line through prognostic index values of patients with normal FHR traces. (B) Cut-off level positioned 1 SD above regression line A. (C) Cut-off level positioned 2 SD above regresion line A.

In contrast, as long as the prognostic index values remained situated below the +2SD cut-off level, within 4 weeks 50% developed an abnormal FHR trace (curve I, lower panel).

In comparing for each patient the last obtained prognostic index value before delivery with umbilical artery pH, it appeared that both variables were significantly correlated ($r = -0.35$, $p=0.03$; $n=40$). The index also significantly correlated with BE ($r = +0.35$, $p=0.03$; $n=39$).

To validate the derived prognostic index, Doppler data of 25 IUGR cases from a previous cross-sectional study[8] were reviewed. This comparison was permitted as it concerns a similar growth-retarded population. Data from 18 patients, who were not displaying an abnormal FHR trace on the day at which the fetal Doppler recording was obtained, were available for this evaluation. Fifteen out of 18 patients subsequently developed an abnormal FHR trace, all but one of which displayed a prognostic index value above both the +1SD and +2SD cut-off level depicted in Figure 1. In the remaining three women, the prognostic index was situated below the +1SD level.

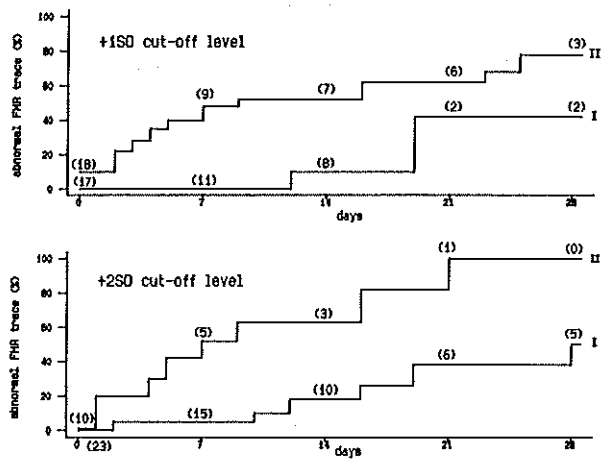


Figure 2. Cumulative percentages of patients developing an abnormal FHR trace. Upper panel shows the results for the +1 SD cut-off level; lower panel refers to the +2 SD cut-off level. For curve I (upper and lower panel) the horizontal axis represents the number of days following the day of entry into the study during which the prognostic index always remained below the + 1 SD cut-off level and + 2 SD cut-off level, respectively. For curve II (upper and lower panel) the horizontal axis represents the number of days following the day that the prognostic index was first situated above the + 1 SD cut-off level and + 2 SD cut-off level, respectively. Numbers between parentheses denote number of patients still at risk of developing an abnormal FHR trace

Discussion

Previous studies in high risk pregnancies have related Doppler flow velocity

waveforms to neonatal outcome, thereby comparing Doppler data with that of fetal heart rate monitoring[5,18]. Unlike these reports, the present study describes the use of Doppler flow measurements as a monitoring tool in predicting abnormal FHR traces. Standardization was achieved by classifying them retrospectively according to the Fischer score which is a comprehensive scoring system.

Our finding of Fischer scores being significantly related to gestational age on the day of entry into the study, is likely to be a consequence of patient selection, since particularly cases of early developing IUGR are referred from district hospitals to academic centres for close fetal monitoring. Taking into account the gestational age, both the PIUA scores and PSVAO scores appeared to have a predictive value regarding the presence of an abnormal FHR trace on the day of entry into the study. However, no additional information is provided by the PSVAO score when the PIUA score is known. The risk of an abnormal FHR trace went up sharply with increasing PIUA.

Of particular interest is the course of pregnancy during hospitalization. Several studies have shown that changes in Doppler parameters may precede deterioration in fetal heart rate patterns[2,16]. This is the first study to demonstrate that PIUA, PIICA, PSVAO and AVAO scores are related to the development of an abnormal FHR trace. Also here, the cardiac parameters do not contribute to the assessment of fetal well-being when the combined PIUA and PIICA scores are known. The derived prognostic index, therefore, only includes the PIUA and PIICA scores. The inclusion of the PIICA emphasises the pathophysiological importance of a reduced cerebral vascular resistance (so-called brain-sparing) in intrauterine growth retardation. Validation of the derived prognostic index in a second group of growth-retarded fetuses confirmed its predictive value.

With the introduction of Doppler flow measurements, a number of reports have appeared regarding its value as a secondary test for identifying the compromised fetus. Lowery et al[13] consider these Doppler recordings to have a significant complementary value to traditional antepartum heart rate testing. However, its value as to the timing of delivery is questioned by others[2]. In this context the time-relationship observed in the present study between the development of abnormal Doppler flow velocity waveforms resulting in a raised prognostic index value and the development of abnormal FHR trace is of particular importance. Nearly 50% of women with a raised prognostic index above the +1SD level developed an abnormal FHR trace within a period of one week, whereas none of the patients who remained below the +1SD level developed abnormal FHR traces within 11 days. For clinical application the +2SD cut-off level may possess a more discriminitive power since 54% of patients with a prognostic index value once above this level develop abnormal FHR traces within 1 week and all patients within 3 weeks. However, 37% of cases developed abnormal FHR traces within 3 weeks although their prognostic index value continuously remained below the +2SD cut-off level (fig.2). The time-relationship we found is much weaker than that reported by Bekedam et al[2]. On the other hand, Arduini et al[1] reported significant changes in peripheral and umbilical vessels close to the onset of late FHR decelerations, but a nadir of cerebral vasodilatation two weeks before the appearance of these abnormal heart rate patterns.

The established correlation between the prognostic index and umbilical cord pH and BE suggests that Doppler flow measurements may reflect fetal well-being. Further prospective studies are needed to obtain guidelines for optimal incorporation of

Doppler measurements in obstetric management. It can be concluded that the PI in the umbilical artery and internal carotid artery are predictive for the developement of an abnormal FHR trace.

References

1. Arduini D, G Rizzo, C Romanini. Changes of Pulsatility Index from fetal vessels preceding the onset of late decelerations in growth-retarded fetuses. Obstet Gynecol 1992; 79: 605-610.
2. Bekedam DJ, GHA Visser, AGJ van der Zee, R Snijders, G Poelmann Weesjes. Abnormal umbilical artery waveform patterns in growth-retarded fetuses: relationship to antepartum late heart rate decelerations and outcome. Early Hum Dev 1990; 24: 79-89.
3. Cox DR. The analysis of binary data. London, Methuen, 1970.
4. Cox DR. Regression models and life tables. J Roy Statist Soc 1972; B34: 187-220.
5. Devoe LD, P Gardner, C Dear, RA Castillo. The diagnostic values of concurrent nonstress testing, amniotic fluid measurement, and Doppler velocimetry in screening a general high-risk population. Am J Obstet Gynecol 1990; 163: 1040-1048.
6. Fischer WM, I Stude, H Brandt (1976): Ein Vorschlag zur Beurteilung des antepartualen Kardiotocogramms. Z Geburtsh Perinat 1976; 180: 117-123.
7. Griffin D, K Bilardo, L Masini, J Diaz-Recasens, JM Pearce, K Willson, S Campbell. Doppler blood flow waveforms in the descending thoracic aorta of the human fetus. Br J Obstet Gynaecol 1984; 91: 997-1006.
8. Groenenberg IAL, JW Wladimiroff, WCJ Hop. Fetal cardiac and peripheral arterial flow velocity waveforms in intrauterine growth retardation. Circulation 1989; 80: 1711-1717.
9. Groenenberg IAL, WCJ Hop, JW Wladimiroff. Doppler flow velocity waveforms in the fetal cardiac outflow tract, reproducibility of waveform recording and analysis. Ultrasound Med Biol 1991; 17: 583-587.
10. Groenenberg IAL, T Stijnen, JW Wladimiroff. Blood flow velocity waveforms in the fetal cardiac outflow tract as a measure of fetal well-being in intrauterine growth retardation. Pediatr Res 1990; 27: 379-382.
11. Kaplan EL, P Meier. Nonparametric estimation from incomplete observations. J Am Stat Assoc 1958; 53: 457-481.
12. Kloosterman GJ. On intrauterine growth. Int J Gynaecol Obstet 1970; 8: 895-912.
13. Lowery CL, BV Henson, J Wan, CG Brumfield. A comparison between umbilical artery velocimetry and standard antepartum surveillance in hospitalized high-risk patients. Am J Obstet Gyencol 1990; 162: 710-714.
14. Marsal K, P Persson. Ultrasonic measurement of fetal blood velocity waveform as a secondary diagnostic test in screening for intrauterine growth retardation. J Clin Ultrasound 1988; 16: 239-244.
15. Nijhuis JG, HFR Prechtl, CB Martin, RSGM Bots. Are there behavioural states in the human fetus? Early Hum Dev 1982; 6: 177-195.
16. Reuwer PJHM, EA Sijmons, GW Rietman, MWM van Tiel, HW Bruinse. Intrauterine growth retardation prediction of perinatal distress by Doppler ultrasound. Lancet 1987; i: 415-419.
17. Trudinger BJ, WB Giles, CM Cook, J Bombardieri, L Collins. Fetal umbilical artery flow velocity waveforms and placental resistance: clinical significance. Br J Obstet Gynaecol 1985; 92: 23-30.
18. Trudinger BJ, CH Cook, L Jones, WB Giles. Comparison of fetal heart rate monitoring and umbilical artery waveforms in the recognition of fetal compromise. Br J Obstet Gynaecol 1986; 93: 171-175.
19. Wladimiroff JW, HM Tonge, PA Stewart. Doppler ultrasound assessment of cerebral blood flow in the human fetus. Br J Obstet Gynaecol 1986; 63: 471-475.
20. Wladimiroff JW, MJ Noordam, JAGW van den Wijngaard, WCJ Hop. Fetal internal carotid artery and umbilical artery blood flow velocity waveforms as a measure of fetal well-being in intrauterine growth retardation. Pediatr Res 1988; 24: 609-612.

CHAPTER 11

PROGNOSIS AND PROSTATIC VOLUME CHANGES DURING ENDOCRINE MANAGEMENT OF PROSTATE CANCER: A LONGITUDINAL STUDY

This chapter was published in Journal of Urology 1992; 147: 962-966 by Z.W. Sneller, W.C.J. Hop, P.J. Carpentier, F.H. Schröder.

Introduction

Endocrine therapy is the standard palliative treatment of prostate cancer at least in those patients with symptomatic metastatic disease. Although initial objective response occurs in about 40% of all patients, which is associated with a symptomatic response in 70 to 80% of all cases, relapse to the hormone independent state of this disease is inevitable.[1] Few parameters exist that allow the prediction of progression under endocrine treatment. In previous reports we have shown that the degree of prostatic volume reduction under endocrine treatment after a 3-month period correlated well with the occurrence of progressive disease.[2-4] If there was a volume reduction of 50% or more early distant progression was unlikely. Another purpose of these early studies was to identify patients with local progression, and to investigate the relationship between progression of the primary tumor and the incidence of distant progression. The previous report was hampered by a relatively small number of patients and an observation period that was limited to 24 months. During the subsequent years the prospective study was continued, thereby increasing the period of observation and recruiting more patients under various forms of endocrine management. The accumulated material of 102 patients with an average followup of 3 years warrants a new review and report. The purpose of the present analysis is, while considering other prognostic factors, to reassess the relationship between prognosis and early (at 3 and 6 months after start of treatment) volume changes of the primary tumor, and to investigate the correlation between progression of the primary tumor and distant progression.

Materials and methods

This study was initiated in May 1979 and recruitment of patients continued through December 31, 1987. Patients were included in this prospective study if they had histopathologically verified adenocarcinoma of the prostate, were not eligible for potentially curative treatment and presented with an indication for endocrine management. Furthermore, patients had to give consent to the study protocol, including repeated transrectal ultrasound examinations, and they had to be previously untreated. A total of 42 patients was not eligible to participate in this study, in addition to those who had undergone previous transurethral resection. Ultrasonography on day 0 was not available in 7 patients, crucial ultrasound studies at 3 and 6 months were missing in 15 and ultrasound studies were not sufficiently evaluable in 4. The follow-up period was too short for clinical evaluation in 3 cases. Six patients had to be treated by other

means shortly after entering the study, including radical prostatectomy (3), transurethral resection of the prostate (2) and radiotherapy to the primary tumor before a 3 month ultrasound was performed (1). Seven patients were omitted from the study for other reasons.

Of the 102 patients 58 were treated by intracapsular orchiectomy immediately after the initial evaluation was completed. Most cases in the M0N+ category had undergone staging lymphadenectomy. The remaining 44 patients were treated with luteinizing hormone-releasing hormone in 3 different forms of application, including 14 who received cyproterone acetate during the first months of therapy. These forms of treatment are generally considered equivalent. Patients were seen every 3 months. After longer periods of stability of the disease process follow-up was extended to 6 months. Follow-up studies included transrectal ultrasonography, markers available at various times and other routine laboratory studies. Bone scans and bone x-rays were done yearly of when indicated. Ultrasound findings had no influence on installing other diagnostic procedures.

Patients left the ultrasound protocol when distant progression or death occurred. Local progression was not a reason to discontinue the study. The criteria of the European Organization for Research on Treatment of Cancer Genitourinary group were used to determine progression.[5] Death from prostatic cancer was defined as death in the presence of a heavy metastatic load with no other obvious cause. Local progression was considered to have occurred when a series of 3 consecutive measurements showed an increase of at least 20% in volume after which the increase was noted. The time of local progression of the primary tumor was considered the time of the second measurement in the consecutive series.

During the initial period of this study a 3.5 MHz. rotating ultrasound probe was used, which was replaced in 1983 by a 7.5 MHz. probe. Both probes were mounted on the Aloka chair allowing axial transverse sections at fixed distances. A specially trained nurse performed all measurements. Prostatic volumes were calculated as described previously.[3] The surface of transverse sections taken at 1 cm. distances was calculated on a planimeter and prostatic volumes were calculated by totalling the surfaces of all transverse sections.

Table 1 gives the tumors, nodes and metastases categories (International Union Against Cancer 1982), age, pretreatment prostate volume and grade according to primary treatment. The grade of differentiation of the primary tumor was determined by review of routine pathology reports according to the system of Mostofi.[6] In case of pleomorphism the worst formation was judged as being representative and was included as a parameter in this study. All cases in the M0 group except for 3 T4N0M0 cases had lymph node metastases.

Survival and time until distant progression were calculated on an actuarial basis using Kaplan-Meier curves. Comparison of these curves was done by logrank tests. Intercurrent death corrected survival was calculated by considering patients withdrawn from study at the moment patients died of causes not related to prostatic carcinoma. The same applied to the calculation of survival without distant progression. Multivariate analysis was done using Cox regression.[7] This method with a time-dependent covariate was applied to evaluate the prognostic value of local progression with respect to distant progression.[8] P values are 2-sided, with 0.05 taken as the limit of statistical significance.

Table 1. Various characteristics at entry into the study according to primary treatment. Data given are numbers of patients or medians (with 10 to 90 percentile range).

		Treatment		
		Luteinizing Hormone-Releasing Hormone	Orchiectomy	Total
T category:	T0-2	10	13	23
	T3	19	26	45
	T4	15	16	31
	TX	0	3	3
N category:	N0	7	18	25
	N1	4	13	17
	N2	9	6	15
	N3	4	6	10
	N4	14	4	18
	NX	6	11	17
M category:	M0	11	24	35
	M1	33	34	67
Grade:	G1	2	3	5
	G2	22	28	50
	G3	19	23	42
	GX	1	4	5
Prostate Vol (cm^3):		29 (17-50)	34 (22-59)	33 (19-58)
Age (yrs):		67 (53-77)	69 (56-79)	68 (54-78)

Results

Of the 102 patients 70 patients died during the study period, including 53 of prostate cancer at an average of 28.7 months (range 3 to 78) after initiation of treatment and 17 of other causes after 34.2 months (range 8 to 84). The remaining 32 patients are alive at the end of the study with a follow-up period of 47.1 months (range 12 to 92). The follow-up period of all 102 patients averaged 53.2 months (range 3 to 92). During the study period distant progression developed in 61 patients. Time to distant progression averaged 18.0 months (range 1 to 73). Distant progression was noted during year 1 in 28 cases, year 2 in 20, year 3 in 6 and thereafter in 7. Median survival, median intercurrent death corrected survival and the median period until distant progression, when calculated on an actuarial basis, were 34, 52 and 24 months, respectively.

Not counting occasional measurements at month 1 and 2 , a total of 594 ultrasound studies was performed. At 3 months the prostatic volume had decreased to a mean of 58% of the pretreatment volume in the orchiectomy group. The mean decrease was significantly (Mann-Whitney's test, $p = 0.02$) greater compared to the luteinizing

hormone-releasing hormone treated group (mean percentage of pretreatment volume 69%). At 6 months these figures were 56% and 64%, respectively (p = 0.08). It is noteworthy that only 2 patients did not show a volume decrease at 6 months, while 99 did (97.1%). A volume decrease of at least 50% at 3 months was noted in 33% of the orchiectomy cases compared to 21% of the luteinizing hormone-releasing hormone group (Fisher's exact test, p = 0.19).

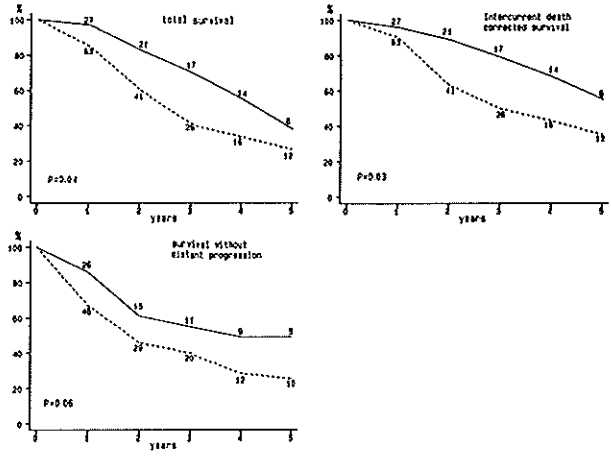


Figure 1. Total survival, intercurrent death corrected survival and survival without progression for patients whose prostate had regressed by at least 50% at 3 months (solid curve) or less than 50 % (dotted curve). Curves are obtained by yearly interpolation of Kaplan-Meier estimates. Numbers along curves denote numbers of patients at risk.

At 6 months these percentages were 47% and 31% (p = 0.18), respectively. Grade was analyzed separately in the M0 group: Gx in 1 case, G1 in 2, G2 18 and G3 in 14. The volume reductions in the 3 groups amounted to 38.0%, 40.6% and 34.4%, respectively, and did not differ statistically. Considering the small numbers and similar volume reductions at 3 months a separate analysis of time to progression per subgroup was not considered useful.

Figure 1 shows survival, intercurrent death corrected survival and survival without distant progression according to whether the prostate had regressed to at least 50% of the pretreatment volume at month 3. Survival was better in cases of such a large decrease. This finding applied also to the distant progression rate, although it is just above statistical significance (p = 0.06). Similar results were obtained when patients were grouped according to whether the volume decrease was at least 50% at month 6. However, multivariate analysis of each end point showed that a volume reduction of 50% or more was not of great importance as a prognosticator when the degree of differentiation and M category were considered. This findings applied to the 3 and 6-month ultrasound studies. Table 2 gives outcomes of the multivariate analysis with

respect to the ultrasound measurement at month 3. From both ultrasound examinations no prognostic value could be found by multivariate analysis when other groupings of the measured volume were analyzed. Also age, T and N categories were not significantly related to survival and distant progression. The same applied to the type of treatment.

Table 2. Multivariate analysis of endpoints survival, intercurrent death corrected survival and distant progression. Data given are relative death, c.q. relative progression rates. Number of patients 95 (missing data: volume reduction 2, grade 5).

Factor	Death (all causes)	Death (Ca)	Distant progression
Grade:			
G1-2	1	1	1
G3	2.5 ($p<.001$)	2.3 ($p<.01$)	1.8 ($p<.05$)
M category:			
M0	1	1	1
M1	3.0 ($p<.001$)	2.9 ($p<.001$)	2.7 ($p<.001$)
Vol. reduction[#]:			
50% or greater	1	1	1
Less than 50%	1.3($p=.3$)	1.5 ($p=.2$)	1.3 ($p=.3$)

(#) Ultrasound finding at month 3.

Survival and distant progression rates according to the 2 most important prognostic factors in this study are given in figures 2 and 3.
According to the ultrasound findings, regrowth of the primary tumor was observed during the follow-up period in 10 (luteinizing hormone-releasing hormone) and in 14 orchiectomy cases. Regular follow-up ultrasound examinations were done in all patients before distant progression. The interval from treatment to regrowth for these cases varied from 3 months (a case that did not show regression of the primary tumor after treatment) to 60 months (median 18 months). In 15 of the 24 patients distant progression was observed at a median interval after the regrowth of 5 months (range 0 to 30 months). Of the 102 patients, 57 were still in follow-up at 18 months without distant progression, including 10 with local progression at or before month 18, and 47 without progression. After month 18 subsequent distant progression was noted in 7 of the 10 and in 17 of the 47 patients. The actuarial cumulative percentage of cases having distant progression after month 18 was significantly ($p<0.01$) greater in the 10 patients compared to the 47 patients. After 2 years these percentages were 77% and 24%, respectively. Using Cox regression, local regrowth correlated significantly with the rate of distant progression, and after regrowth the risk of distant progression was

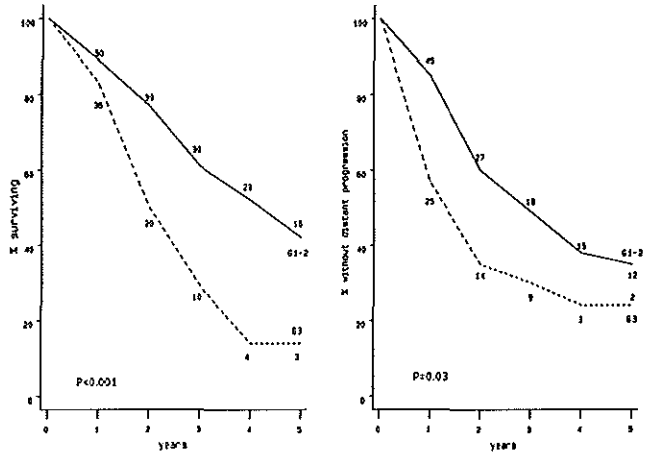


Figure 2. Overall survival and progression-free survival according to G category.

2.7 times higher ($p<0.01$).[8]
Local regrowth also appeared to be a prognostic factor for distant progression when the degree of differentiation and M category were considered (table 3).

Table 3. Multivariate analysis of distant progression according to grade, M category and local progression. Data given are relative rates of distant progression.

Factor		Relative rate	Significance
Grade:	G1-2	1	
	G3	1.7	$p=.06$
M category:	M0	1	
	M1	2.4	$p<.005$
Local Progression:	No	1	
	Yes	2.4	$p<.01$

Discussion

This longitudinal study of 102 patients with prostate cancer confirms only partially the previous findings of a significant correlation between the change of overall prostatic volume and its value in predicting progressive disease. In the previous study, which did not include a multivariate analysis, prostatic volume reduction under endocrine treatment was described as being a prognostic factor. Although a volume reduction of at least 50% at 3 months in this study was significantly related to survival (all causes

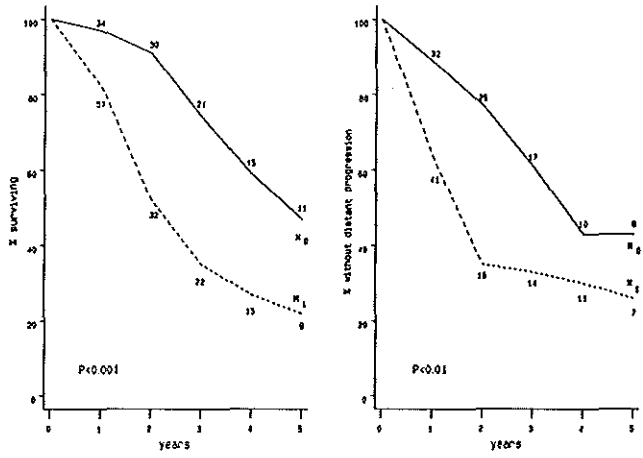


Figure 3. Overall survival and progression-free survival according to M-category.

and death from prostate carcinoma) and almost significantly to distant progression, the value of this observation is limited by the fact that grade of malignancy and M status together are of overruling prognostic significance. After adjustment for both of these factors, no significant relation between the volume reduction and prognosis was present, which could be explained by the findings that there were relatively more patients with a favourable grade and M category in the group with a greater volume decrease. Of 69 cases with G1-2MO disease 12 (17%) had a volume reduction of less than 50% at month 3, whereas 8 of 26 (31%) had a greater prostate volume reduction.

In the previous report distant progression was seen in 20 of 55 patients during a 24-month interval. At the same time local progression occurred in only 2 instances (10%). In this report with an average follow-up of 53.2 months systemic progression was seen in 61 patients, while local progression occurred in 24.

Although there appears to be a statistical association between local progression and an increased risk of subsequent distant progression, the clinical value of monitoring the prostate volume for this purpose seems slight. Many cases that show local progression do not develop distant progression, while also there are many cases with distant progression without prior local progression. The risk of local volume increase overall cannot reliably be assessed because patients went off protocol at the time of distant progression. Nevertheless 6 patients had shown local progression after distant progression but this was not systematically investigated. Micturition problems requiring treatment were seen in 9 patients (9%) (some with and some without local progression). The fact that 99 of 102 patients showed at least some volume reduction indicates the responsiveness of the primary tumor but it may also be related to associated prostatic hyperplasia. The study by Ryan et al represents the only other report in the literature that has pursued the same approach as we did.[9] They have not been able to find in 84 men treated by subcapsular orchiectomy or with goserelin acetate depot a correlation between prostatic volume reduction and progression observed in 20 patients during a period of 12 months.

Obviously, it would be advantageous to be able to measure prostatic cancer volume rather than prostatic volume. Lee et al,[10] and Frentzel-Beyme and Ledwa[11] clearly showed that many if not most prostate cancers can be identified as hypoechogenic areas within the remainder of the prostate gland. However, Rifkin et al believe that a substantial proportion of prostate carcinomas present as hyperechogenic and isoechogenic lesions.[12] When Carter et al used transrectal ultrasonography studies that were obtained before radical prostatectomy to identify prostate cancer in apparently normal contralateral lobes of cancerous prostates in comparison to the histological findings, a sensitivity and specificity of only 52% and 68% resulted.[13] Especially the low specificity indicating the large proportion of false positive ultrasound studies makes it difficult to follow hypoechogenic lesions as a parameter for volume change unless these individual lesions are identified by biopsy as representing carcinoma of the prostate.
Considering the inaccuracy of digital rectal examination in identifying correctly the volume of the prostate or the volume of prostatic cancer within the gland, it seems reasonable to obtain volume determinations before treatment and again during treatment as soon as there is suspicion of a prostatic volume increase or if serum markers increase. A volume increase of more than 20% in our study seems to indicate early which patients tend to have a higher risk for distant progression.
Early transurethral resection even for palliative reasons or palliative radiotherapy may be applicable to such patients to prevent the difficulties arising from a large pelvic tumor.
Of the 102 patients 58 were treated by castration, 30 with luteinizing hormone-releasing hormone alone and 14 with luteinizing hormone releasing hormone plus 150 mg. cyproterone acetate daily for 1 month. When the prostatic volume reduction seen at 3 months in the 58 castrated patients was compared to the volume change in the remaining endocrine treatment group a significant difference in favour of a more pronounced volume reduction in the castration group was seen. Also, the only 2 patients in whom no volume decrease was observed after treatment had received luteinizing hormone-releasing hormone. At 6 months and thereafter no significant differences in mean prostate volume were present. This finding suggests that it takes longer with luteinizing hormone-releasing hormone treatment to reach a maximum response of the target cells and that in this respect castration may be superior. The ultrasound studies were well tolerated and accepted by the patients. No side effects occurred and few refusals of further studies were encountered.

References

1. Pavone-Macaluso M, HJ de Voogt, G Viggiano, E Barasolo, B Lardennois, M de Pauw, R Sylvester. Comparison of diethylstilbestrol, cyproterone acetate and medroxyprogesterone acetate in the treatment of advanced prostatic cancer: final analysis of a randomized phase III trial of the European Organization for Research on Treatment of Cancer Urological Group. J Urol 1986; 136: 624.
2. Carpentier PJ, FH Schröder. Transrectal ultrasonography in the following of prostatic carcinoma patients: a new prognostic parameter? J Urol 1984; 131: 903.
3. Carpentier PJ, FH Schröder, PIM Schmitz. Transrectal ultrasonometry of the prostate: the prognostic relevance of volume changes under endocrine management. World J Urol 1986; 4: 159.
4. Carpentier PJ, FH Schröder, JHM Blom. Transrectal ultrasonography int he follow-up of prostatic

carcinoma patients. J Urol 1982; 128: 742.
5. Schröder FH. Treatment response criteria for prostatic cancer. Prostate, 5:181,1984.
6. Mostofi FK. Grading of prostatic carcinoma. Cancer Chemother Rep 1975; 59: 111.
7. Cox DR. Regression models and life-tables. J Roy Stat Soc 1972; B34: 187.
8. Hop WCJ, HR van Buuren. A method to evaluate changes in prognostic status during follow-up. Comp Biol Med 1989; 19: 181.
9. Ryan PG, DR Jones, WB Peeling, EE Roberts. Transrectal ultrasonography and CT scanning of the primary tumour in the follow-up of patients with prostatic carcinoma. In: EORTC Genitourinary Group Monograph 5: Progress and Controversies in Oncological Urology II. New York: Alan R. Liss, Inc., pp. 57-65, 1988.
10. Lee F, JM Gray, RD McLeary, TR Meadows, GH Kumasaka, GS Borlaza, WH Straub, F Lee, MH Solomon, TA McHugh, RM Wolf. Transrectal ultrasound in the diagnosis of prostate cancer: location, echogenicity, histopathology, and staging. Prostate 1985; 7: 117.
11. Frentzel-Beyme B, D Ledwa. Die Echomorphologie der Prostata. Ultraschall Med 1986; 7: 7.
12. Rifkin MD, ET McGlynn, H Choi. Echogenenicity of prostate cancer correlated with histologic grade and stromal fibrosis: endorectal US studies. Radiology 1989; 170: 549.
13. Carter HB, UM Hamper, S Sheth, RC Sanders, JI Epstein, PC Walsh. Evaluation of transrectal ultrasound in the early detection of prostate cancer. J Urol 1989; 142: 1008.

CHAPTER 12

GENERAL DISCUSSION

Diseases develop in time and the same will apply to their progress. Important aspects in the characterization of a particular disease are the effects of therapy and the prognosis of the patients. In view of the vast growing amount of literature, these aspects receive much attention. A search in the literature database MEDLINE, covering the period 1992 up to mid 1996, resulted in a number of more than 25000 articles in which the keywords prognosis or prognostic had been used. For breast cancer, a disease with a high incidence, Klein et al (1993) have compiled a list of prognostic factors, mostly of biochemical origin, which amounts to about one hundred. Also in the statistical litterature there has been a growth of papers dealing with the statistical methodology of the subject. A milestone was the pioneering paper of Cox (1972) in which the innovation of the proportional hazards model was described. Besides the more elementary methods such as the Kaplan-Meier curve and the logrank-test, the proportional hazards model is probably the most widely used statistical tool in investigations regarding the prognostic value of patient characteristics and the effects of therapy (it can be siad that Kaplan-Meier curves and logrank-tests can even be considered to be special elaborations of the Cox-regression method). The method has gained great popularity to evaluate baseline charateristics. Due to its flexibility, it may also be extremely valuable in deriving rules for updating prognosis for patients according to the history observed during the course of the disease.
To sort out the most important prognostic factors, usually a step-wise approach is chosen. The most logical option is to perform a backward elimination approach by eliminating factors from the model which have the least significant contribution to the model (forward selection is less appropriate due to the fact that the model cannot fit fully in the initial stages of the building of a model). In this process, however, the correlation structure of the various prognostic factors has to be taken into account. If two variables are strongly correlated, it will be difficult to assess their separate prognostic value. Also factors may need to be retained because of prior knowledge or because there is special interest in the factor possibly due to its putative causal effect. The selection procedure will result in a predictive rule which will tend to perform less when it is applied to a new set of patients as compared to the performance of the rule when it is applied to the group of patients from which the prognostic index was derived (Phillips et al, 1990). With a very large dataset, the effect of this phenomenon can be investigated by randomly splitting the dataset in two halves. One half is used to determine the prognostic index, while the other half is used for verification of the prognostic performance of the rule. In practice, however, this is not a realistic option due to the large number of patients required. Using the data on which the predictive rule is based, Verweij and Van Houwelingen (1993) have proposed the method of cross-validation to assess the ability of the proportional hazards model to predict future data. The method of choice, however, is the testing of the clinical prediction rule in another group of similar patients. These new patients can come from the clinical setting from which the prognostic index was derived, but a different clinical environment is to be preferred (Wasson et al, 1985). Such a prospective validation generally will be costly and time consuming, but it is essential in order to obtain a

wider acceptancy of any prognostic index.
Sometimes there are treatments for life threatening diseases for which there appears to be so much evidence about their efficacy from uncontrolled studies, that it is considered not clinically valid to initiate randomized studies to prove their efficacy. Heart and liver transplantations are examples. Cox-regression with time-dependent variables, using only concurrent controls from the waiting list for transplantation and allowing for important prognostic determinants, is an option to evaluate the efficacy of such transplantations. Another option which remains in such circumstances is to prove that the outcome for patients treated that way is better than the outcome for similar patients before the introduction of the treatment. Bonsel et al (1990) thus compared the survival of patients with primary biliary cirrhosis after transplantation with expected survival, the latter being determined according to various published prognostic indices which all had been derived using Cox-regression.
The successful application of Cox-regression in many publications is no guarantee that it will work satisfactorily in all situations. The information contents of any dataset are limited. It is impossible to assess the independent association of factors with prognosis from material in which the number of potential predictors approaches the number of patients studied. This applies to Cox-regression as well as to any other statistical method. The number of patients (actually, the number of events present in the dataset) should be large in comparison with the number of factors studied. It is advocated, as a rule of thumb, that a minimum of 10 patients who experienced the event are required for each factor studied [Harrell et al, 1985]. (There are good reasons to count a categorical factor with k levels even as a number of k-1 factors in this calculation). No thoroughly based rules for the number of patients required, however, are available in multivariate situations. The difficulty in deriving such rules is caused by the fact that the correlation structure of the various predictors need to be known, which will not be the case.
Publications in which the prognostic effect is evaluated of factors as observed during follow-up are relatively rare, in contrast to evaluations of baseline characteristics. In publications of the first type, however, often inappropriate methods have been used. There seems to be much room for improvement here, although the situation improves in the last few years.
There are occasions in which time does not play an important role in evaluating prognosis, e.g. in assessing which factors are predictive for a successful operation. For such short-term prognostic evaluations, logistic regression has become the preferred method [Cox, 1970].
In most situations, however, the method of Cox-regression has become the method of choice. It is likely that the method will keep that status for a very long time.

References

Bonsel GJ, IJ Klompmaker, F van 't Veer, JDF Habbema, MJH Slooff. Use of prognostic models for assessment of value of liver transplantation in primary biliary cirrhosis. Lancet 1990; 335: 493-497.

Cox DR. The analysis of binary data. 1970. Methuan, London.

Cox DR. Regression models and life tables (with discussion). J Roy Stat Soc 1972; B34: 187-220.

Harrell FE, KL Lee, DB Matchar, TA Reichert. Regression models for prognostic prediction: advantages, problems, and suggested solutions. Cancer Treatm Rep 1985; 69: 1071-1077.

Phillips AN, SG Thompson, SJ Pocock. Prognostic scores for detecting a high risk group: estimating the sensitivity when applied to new data. Stat Med 1990; 9: 1189-1198.

Klijn JGM, EMJJ Bems, JA Foekens. Prognostic factors and response to therapy in breast cancer. Cancer Surveys 1993; 18: 165-198.

Verweij PJM, JC van Houwelingen. Cross-validation in survival analysis. Statist Medic 1993; 12: 2305-2314.

Wasson JH, HC Sox, RK Neff, L Goldman. Clinical prediction rules; applications and methodological standards. New Eng J Med 1985; 313: 793-799.

Summary

The various roles which knowledge about prognostic factors may play in clinical medicine are discussed in Chapter 1. Some well-known prognostic staging systems for a few specific diseases are described briefly.

Some important theoretical statistical developments of what is usually referred to as Survival Analysis are reviewed in Chapter 2. Emphasis is placed on the concept of Hazard Rate Function. In many instances this function is better suited to assess the role of patient or disease characteristics and/or interventions regarding their impact on prognosis than the more traditional Survival Function. The flexibility of the method of Cox-regression, by which the prognostic value of various factors can be assessed simultaneously, is emphasized.

A case study in patients with kidney carcinoma is presented in Chapter 3. Using Cox-regression the role of various baseline factors regarding their impact on survival prognosis of operated patients was evaluated. Besides gender and tumour invasion of the renal vein, it was found that a very low-cost parameter, namely the level of the erythrocyte sedimentation rate, was an important indicator of prognosis.

Cox-regression played an important role in the analysis of possible adverse effects of blood transfusions regarding prognosis in patients operated for colorectal cancer as described in Chapter 4. The main finding in a randomized clinical trial was that there were no significant differences regarding recurrence rates and survival between patients who had donated blood before the operation and those who had to receive allogeneic blood if required. It was also found that patients who needed blood transfusions had a worse prognosis as compared to those who did not need transfusions. This applied equally to both types of blood transfused.

Another case study is described in Chapter 5. Various prognostic factors were evaluated using Cox-regression regarding their impact on prognosis in patients operated for esophageal carcinoma. The calculation of the life-time risk of dysphagia due to recurrent disease is described in an appendix to this chapter.

Chapter 6 discusses some pitfalls regarding the assessment of the prognostic value of factors whose level may change during follow-up. Theoretical developments of various relevant methods, among which Cox-regression with time-dependent variables, to perform such assessments reliably are described.

In Chapter 7 a case study is presented which shows an analysis using time-dependent covariates. It was found that the occurrence of a Gram-negative infected pancreatic necrosis in patients with severe acute pancreatitis is a very severe ominous sign. Using Cox-regression, with type of infection of the necrosis as a time-dependent factor, it turned out that the observed smaller mortality using Adjuvant Selective Decontamination could be explained by a reduced incidence of such Gram-negative infections.

Chapter 8 gives an account of an investigation of factors related to the risk of a spontaneous abortion among pregnant women scheduled for transabdominal chorionic

villus sampling. The risk of the procedure itself was assessed by considering patients who were still scheduled for the test to act as controls for those women who already underwent the procedure. An increased risk could not be demonstrated.

The prognostic value of cytoimmunological monitoring of peripheral blood from patients who had received a heart transplant was investigated using Cox-regression with time-dependent variables as described in Chapter 9. The critical event considered was the first occurrence of a rejection episode. It is concluded that the method is not able to replace the currently required invasive method of taking a biopsy.

In Chapter 10 the monitoring value of various Doppler measurements in growth retarded fetusses in relation to the occurrence of abnormal Fetal Heart Rate traces is discussed. Using Cox-regression with the time-dependent serial Doppler measurements, it was possible to derive a prognostic index regarding the risk of abnormal fertal heart rate traces.

In Chapter 11 the prognostic value of serially performed ultrasound measurements of the prostate, in addition to other factors, was investigated in prostatic cancer patients who received endocrine treatment.

A general discussion of the statistical methodology of the assessment of prognostic factors in clinical medicine is presented in Chapter 12.

Samenvatting

De verschillende wijzen waarop kennis omtrent prognostische factoren in de klinische geneeskunde van belang kan zijn worden besproken in Hoofdstuk 1. Enkele bekende prognostische stadiëringssystemen voor enkele specifieke ziekten worden toegelicht.

Enige belangrijke theoretische statistische ontwikkelingen op het gebied van wat gebruikelijk wordt aangeduid met Overlevingsanalyse worden besproken in Hoofdstuk 2. De nadruk is gelegd op het concept Uitvalstempofunctie (Hazard Rate Function). In vele gevallen is deze functie beter geschikt om het belang van patiënten- of ziektekarakteristieken en/of interventies na te gaan dan de meer gebruikelijke Overlevingsfunctie. De flexibiliteit van Cox-regressie, door middel waarvan de prognostische waarde van meerdere factoren kan worden nagegaan, wordt benadrukt.

Een studie over patiënten met nierkanker wordt gepresenteerd in Hoofdstuk 3. Gebruikmakend van Cox-regressie werd het belang van diverse uitgangsvariabelen met betrekking tot de overleving van geopereerde patiënten nagegaan. Naast geslacht en het gegeven of de vena renalis was aangetast door de tumor, bleek het niveau van de bloedbezinking een belangrijke indicator voor de prognose te zijn.

Cox-regressie speelde een belangrijke rol bij de analyse van mogelijke nadelige effecten van bloedtransfusies met betrekking tot de prognose bij patiënten die geopereerd werden voor dikke darm of rectum kanker zoals beschreven in Hoofdstuk 4. De voornaamste bevinding in een gerandomiseerde klinische studie was dat er geen significante verschillen waren wat betreft recidiefkansen en overleving tussen patiënten die voor de operatie bloed hadden afgestaan en patiënten die op de gebruikelijke wijze allogeen bloed kregen indien nodig. Er werd ook gevonden dat patiënten die bloedtransfusies nodig hadden een slechtere prognose hebben dan zij die geen bloedtransfusies nodig hadden. Dit gold in gelijke mate voor beide typen bloedtransfusies.

Een andere studie is beschreven in Hoofdstuk 5. Verschillende variabelen werden geëvalueerd met Cox-regressie betreffende hun prognostisch belang bij patiënten die geopereerd waren voor slokdarmkanker. De berekening van het absolute risico op slikklachten ten gevolge van een recidief van de ziekte is besproken in een appendix van dit hoofdstuk.

Hoofdstuk 6 bespreekt enkele voetangels en klemmen met betrekking tot de evaluatie van factoren waarvan het niveau kan veranderen gedurende de follow-up. Theoretische ontwikkelingen van diverse relevante methoden, waaronder Cox-regressie met tijds-afhankelijke variabelen, om dergelijke evaluaties betrouwbaar uit te kunnen voeren, worden besproken.

In Hoofdstuk 7 wordt een studie gepresenteerd welke een analyse toont gebruikmakend van tijds-afhankelijke variabelen. Er werd gevonden dat het optreden van gramnegatieve infecties van alvleesklierneerose bij patiënten met acute pancreatitis een ongunstig voorteken is. Met Cox-regressie, waarbij type infectie van de alvleesklierneerose de rol van tijds-afhankelijke variabele vervulde, bleek dat de gevonden lagere

mortaliteit bij het gebruik van Selectieve Darm Decontaminatie verklaard kan worden door een lagere incidentie van zulke gram-negatieve infecties.

Hoofdstuk 8 geeft een bespreking van een onderzoek naar factoren die gerelateerd zijn aan het risico op spontane abortus onder zwangere vrouwen die een transabdominale vlokkentest zouden ondergaan. Het risico van de procedure zelf werd nagegaan door die vrouwen die een afspraak hadden staan voor het onderzoek te beschouwen als controlepatiënten voor diegenen die de test reeds ondergaan hadden. Een verhoogd risico ten gevolge van de procedure kon niet worden aangetoond.

De prognostische waarde van cytoimmunologische testen van perifeer bloed bij patiënten die een harttransplantatie hadden ondergaan werd met behulp van Cox-regressie met tijdsafhankelijke variabelen onderzocht zoals beschreven in Hoofdstuk 9. De conclusie luidt dat de methode niet geschikt is om als vervanging te dienen voor de huidige invasieve methode van het nemen van biopten uit het hart om een afstoting te kunnen vaststellen.

In Hoofdstuk 10 wordt de waarde van diverse Doppler metingen bij groeivertraagde foetussen onderzocht wat betreft hun vermogen om het optreden van foetale nood te voorspellen. Door middel van Cox-regressie, met de seriële Doppler metingen als tijds-afhankelijke variabelen, was het mogelijk een prognostische index te ontwikkelen met betrekking tot het optreden van foetale nood.

In Hoofdstuk 11 werd, additioneel op andere factoren, de prognostische waarde van seriële echografische metingen van de prostaat onderzocht bij prostaatkanker patiënten die hormonaal behandeld werden.

Een algemene beschouwing van de statistische methoden die relevant zijn voor het evalueren van prognostische factoren in de klinische geneeskunde wordt gegeven in Hoofdstuk 12.

Nawoord

Allereerst bedank ik mijn promotor Prof. R. van Strik. Beste Roel, afgezien van een korte onderbreking ben jij voor mij sinds 1974 een grote steun geweest in het werk. Zoals je weet moet, in verband met allerlei deadlines, in de analysefase vaak de tijd ingehaald worden die door tegenslagen in het vóórtraject verloren kan zijn gegaan. Jouw bereidwilligheid om steeds snel tijd beschikbaar te maken voor overleg aangaande ingewikkelde problemen uit de consultatie heb ik dan ook zeer gewaardeerd.
Mede dankzij jouw herhaalde aansporingen is dit boekje verschenen. Dank voor jouw constructieve suggesties bij de samenstelling ervan.

Het verschijnen van dit proefschrift is mogelijk gemaakt met de hulp van velen. Het zou te ver gaan om iedereen met name te noemen. Daarom beperk ik mij, in volgorde van de hoofdstukken, tot de eerste gesprekspartners bij de beschreven studies:
Prof. Dr B.H.P van der Werf-Messing, Dr O.R.C. Busch, Dr J.J. van Lanschot, Drs H. van Buuren, Drs E.J.T. Luiten, Drs T.E. Cohen-Overbeek, Dr N.H.P.M. Jutte, Dr I.A.L. Groenenberg en Drs Z.W. Sneller.
Beste Brigit, Olivier, Jan, Henk, Ernest, Titia, Nicolette, Irene en Wim, als niet-clinicus heb ik de besprekingen over de projecten altijd als zeer interessant en leerzaam ervaren. Ik vond het een voorrecht te mogen meedenken met zulke enthousiaste onderzoekers.

Tenslotte bedank ik Drs J.A. Bakkum voor het nakijken van de Engelse tekst van enkele hoofdstukken. Ook, Soeja de Groot en mijn zoon Walter, bedankt voor jullie geweldige hulp bij het op papier krijgen van de tekst. WP stelt je soms voor onaangename verrassingen, maar dankzij jullie kwam er altijd een oplossing.

Curriculum Vitae

W.C.J. Hop is geboren te Made op 4 april 1947. In 1964 behaalde hij het diploma HBS-B aan het Mgr Frencken College te Oosterhout. Na aanvankelijk Technische Natuurkunde aan de Technische Hogeschool te Delft te hebben gestudeerd, maakte hij in 1967 de overstap naar de studierichting Wiskunde. Hier behaalde hij in 1974 het diploma Wiskundig Ingenieur, met als specialisatie Mathematische Statistiek. Zijn afstudeerstage verrichtte hij bij TNO.
Gedurende de jaren 1974-1982 was hij werkzaam als medisch statisticus bij de Dr Daniel den Hoed Kliniek te Rotterdam. Tevens was hij gedurende die tijd als honorair medewerker verbonden aan het Instituut voor Biostatistica van de Medische Faculteit te Rotterdam. Als docent Wiskunde en Informatica was hij gedurende de periode 1982-1986 verbonden aan de Hogeschool voor Techniek en Gezondheidszorg te Enschede. Sindsdien is hij werkzaam als universitair docent Biostatistiek bij het Instituut voor Epidemiologie en Biostatistiek van de Faculteit der Geneeskunde en Gezondheidswetenschappen van de Erasmusus Universiteit te Rotterdam. Een dag per week vervult hij een statistisch consulentschap bij Glaxo Wellcome te Zeist. Hij is lid van de Commissie Klinisch Vergelijkend Onderzoek (CKVO) van de Nederlandse Kankerbestrijding/KWF. Tevens is hij lid van de Medisch Ethische Commissie van het St Clara Ziekenhuis te Rotterdam.

Lijst van publicaties:

F.A.G Teulings, H.A. Peters, W.C.J. Hop, W. Fokkens, W.G. Haije, H. Portengen, B. van der Werf-Messing. A new aspect of the urinary excretion of tryptophan metabolites in patients with cancer of the bladder. Int J Cancer 1978,21, 140-146

W.C.J. Hop, B.H.P. van der Werf-Messing. Prognostic indexes for renal cell carcinoma. Eur J Cancer 1980; 16: 833-840.

S. Boulis-Wassif, W.C.J. Hop. The role of preoperative adjuvant therapy in management of borderline operability of rectal cancer. Coloproctology 1980; 2: 320-326.

J. Alexieva-Figush, H.A. van Gilse, W.C.J. Hop, C.H. Phoa, J. Blonk-van der Wijst, R.E. Treurniet. Progestin therapy in advanced breast cancer. Megestrol acetate, an evaluation of 160 treated cases. Cancer 1980; 46: 2369-2372.

F.H.J. Rampen, W.A. van Houten, W.C.J. Hop. Incisional procedures and prognosis in malignant melanoma. Clin Exper Dermatology 1980; 5: 313-320.

A.D. Treurniet-Donker, W.C.J. Hop, S. den Hoed-Sijtsema. Radiation treatment of stage III mammary carcinoma. Int J Radiat Oncol Biol Phys 1980; 6: 1477-1482.

B. van der Werf-Messing, W.C.J. Hop. Carcinoma of the urinary bladder treated either by Radium implant or by transurethral resection only. Int J Radiat Oncol Biol Phys 1981; 7: 299-303.

K. Sintnicolaas, H.M. Vriesendorp, W. Sizoo, W.F. Stenfert-Kroese, W.G. Haije, W.C.J. Hop, J. Abels, B. Löwenberg. Delayed alloimmunisation by random single donor platelet transfusions. Lancet 1981; april 4: 750-754.

B. van der Werf-Messing, W.C.J. Hop. Non-seminomateuze kiemceltumoren van de testis. Ned T Geneesk 1981, 125: 1281-1286.

W.C.J. Hop, J. Hermans. Statistische analyse van overlevingsduren. T Soc Geneesk 1981; 59: 279-288.

B. van der Werf-Messing, W.C.J. Hop. Blaaskanker, categorie T1NXM0, behandeld met radiumimplantatie of transuretrale resectie. Ned T Geneesk 1981, 125: 825-829.

J.G. van Andel, W.C.J. Hop. Carcinoma of the nasopharynx, a review of 86 cases. Clin Radiol 1982; 33: 95-99.

W.H. Bouma, J.H. Mulder, W.C.J. Hop. The influence of intramedullary nailing on the development of metastases in the treatment of an impending pathological fracture. Neth J Surg 1982;34: 159-162.

B.H.P. van der Werf-Messing, G.H. Friedell, R.S. Menon, W.C.J. Hop, S. Boulis Wassif. Carcinoma of the urinary bladder T3NXM0 treated by preoperative irradiation followed by simple cystectomy. Int J Radiat Oncol Biol Phys 1982;8: 1849-1855.

P.J. Carpentier, P.J. Bettink, W.C.J. Hop, F.H. Schröder. Reflux, a retrospective study of 100 ureteric reimplantations by the Politano-Leadbetter method and 100 by the Cohen technique. Brit J Urol 1982;54: 230-233.

J.C.J. Wereldsma, W.C.J. Hop. Resultaten van 10 jaar curatieve chirurgie van colon- en rectumcarcinoom. Ned T Geneesk 1982, 126: 1262-1268.

B. van der Werf-Messing, W. Hop. Radiation therapy of testicular non-seminimas. Int J Radiat Oncol Biol Phys 1982;8: 175-178.

B. van der Werf-Messing, R.S. Mennon, W.C.J. Hop. Carcinoma of the urinary bladder category T3NXM0 treated by the combination of Radium implant and external irradiation; second report. Int J Radiat Oncol Biol Phys 1983;9: 177-180.

R.J.L. Caspers, W.C.J. Hop. Radiotherapy of true pelvis for bladder and prostatic carcinoma in supine, prone or Trendelenburg position. Int J Radiat Oncol Biol Phys 1983;9: 589-593.

K.H. Kurth, P.A. Maksimovic, W.C.J. Hop, F. Schröder, N.J. Bakker. Single dose intravesical Epodyl after TUR of Ta TCC bladder carcinoma. World J Urol 1983;1: 89-93.

G. Baris, J.G. van Andel, W.C.J. Hop. Carcinoma of the tonsillar region. Strahlentherapy 1983;159: 138-142.

B. van der Werf-Messing, R.S. Menon, W.C.J. Hop. Cancer of the urinary bladder category T2, T3 (NXM0) treated by interstitial Radium implant; second report. Int J Radiat Oncol Biol Phys 1983;9: 481-485.

F.H. Schröder, J.H.M. Blom, W.C.J. Hop, F.K. Mostofi. Incidental carcinoma of the prostate treated by total prostatectomy. The prognostic impact of microscopic tumor extension and grade. World J Urol 1983;1: 15-23.

W.H. Bouma, J.H. Mulder, W.C.J. Hop. The influence of intramedullary nailing upon the development of metastasis in the treatment of an impending pathological fracture: an experimental study. Clin Exp Metastasis 1983;1: 205-209.

B. van der Werf-Messing, R.S. Menon, W.C.J. Hop. Interstitial radiotherapy of carcinoma of the bladder at the Rotterdam Radiotherapy Institute. Prog Clin Biol Res 1984;162B: 71-85.

J. Alexieva-Figush, M.A. Blankenstein, W.C.J. Hop, J.G.M. Klijn, S.W.J. Lamberts, F.H. de Jong, R. Docter, H. Adlercreutz, H.A. van Gilse. Treatment of metastatic breast cancer patients with different dosages of Megestrol Acetate; dose relations, metabolic and endocrine effects. Eur J Cancer Clin Oncol 1984;20: 33-40.

V. Kuenen-Boumeester, W.C.J. Hop, D.I. Blonk, M.E. Boon. Prognostic scoring using cytomorphometry and lymph node status of patients with breast carcinoma. Eur J Cancer Clin Oncol 1984;20: 337-340.

J. Alexieva-Figush, F.A.G. Teulings, W.C.J. Hop, J. Blonk, H. van Gilse. Steroid receptors in megestrol acetate therapy. Recent Results Cancer Res 1984;91: 65-678.

F.H. Schröder, J.H. Blom, W.C.J. Hop, F.K. Mostofi. Grading of prostatic cancer (I); an analysis of the prognostic significance of single characteristics. The Prostate 1985;6: 81-100.

F.H. Schröder, J.H. Blom, W.C.J. Hop, F.K. Mostofi. Grading of prostatic cancer (II); the prognostic significance of the presence of multiple architectural patterns. The Prostate 1985;6: 403-415.

F.H. Schröder, W.C.J. Hop, J.H. Blom, F.K. Mostofi. Grading of prostatic cancer (III); multivariate analysis of prognostic parameters. The prostate 1985;7: 13-20.

A.M. Horrevorts, J. de Witte, J.E. Degener, G. Dzoljic, W. Hop, O. Driessen, M.F. Michel, K.F. Kerrebijn. Tobramycin in patients with cystic fibrosis. Adjustment in dosing interval for effective treatment. Chest 1987, 92, 844-848.

S.K. Singh, R.L. Marquet, R.W.F. de Bruin, W.C.J. Hop, D.L. Westbroek, J. Jeekel. Consequences of blood loss on growth of artificial metastases. Br J Surg 1988;75: 377-379.

M.C.W. Scholtes, W.C.J. Hop, A. Alberda, H. Janssen-Caspers, R.A. Leerentveld, H.C. van Os, G.H. Zeilmaker. Factors influencing the outcome of successive IVF treatment cycles in attaining a follicular puncture. Human Reprod 1988;3: 755-759.

H.C. Hoogsteden, J.J.M. van Dongen, H.J. Adriaansen, H. Hooijkaas, M. Delahaje, W.C.J. Hop, C. Hilvering. Bronchoalveolar lavage in extrapulmonary sarcoidosis. Chest 1988;94: 115-118.

H.W.J. van Marwijk, F.M. Bekker, W.C.J. Hop, P.A.F. Jansen, J.F. van Nieuwkerk. Elektroconvulsietherapie by depressieve bejaarden; een retrospectief onderzoek naar werkzaamheid en veiligheid. Ned T Geneesk 1988; 132: 1396-1399.

J. W. Wladimiroff, M.J. Noordam, J.A.G.W. van den Wijngaard, W.C.J. Hop. Fetal internal carotid and umbilical artery blood flow velocity waveforms as a measure of fetal well-being in intrauterine growth retardation. Pediatr Res 1988; 5: 609-612.

B.H.P. van der Werf-Messing, R.S. Menon, W.C.J. Hop. Carcinoma of the urinary bladder category T2T3NXM0 treated by interstitial radium implant. Prog Clin Biol Res 1988; 260: 511-524.

B.H.P. van der Werf-Messing, G. Friedell, R.S. Menon, W.C.J. Hop, S.B. Wassif. Carcinoma of the urinary bladder category T3NXM0 treated by preoperative irradiation followed by simple cystectomy. Prog Clin Biol Res 1988; 260: 453-460.

A.A.M. Franken, F.H.M. Derkx, M.A.D.H. Schalenkamp, A.J. Man in 't Veld, W.C.J. Hop, E.H. van Rens, P.T.V.M. de Jong. Association of high plasma prorenin with diabetic retinopathy. J Hypertension 1988; 6 (suppl. 4): 461-463.

N.H.P.M. Jutte, R. Daane, J.M.G. van den Bemd, W.C.J. Hop, C.E. Essed, M.L. Simoons, E. Bos, W. Weimar. Cytoimmunological monitoring to detect rejection after heart transplantation. Transplantation Proceedings 1989; 21: 2519-2520.

H.C. Hoogsteden, J.J.M. van Dongen, P.T.W. van Hal, M. Delahaye, W. Hop, C. Hilvering. Phenotype of blood monocytes and alveolar macrophages in interstitial lung disease. Chest 1989; 95: 574-577.

J. Nauta, W.C.J. Hop, W.F.A. Grose, A. van der Heyden, J.B. Kist, E.D. Wolff. Improved renal transplant monitoring in outpatient clinics. Transplantation 1989; 47: 715-717.

J.A.G.W. van den Wijngaard, J.W. Wladimiroff, I.A.L. Groenenberg, W.C.J. Hop. Cerebral Doppler ultrasound of the human fetus. Brit J Obstret Gynaecol 1989; 96: 845-849.

W.C.J. Hop, H.R. van Buuren. A method to evaluate changes in prognostic status during follow-up, illustrated by an assessment of the effect of a rebleed in patients with oesophageal variceal bleeding. Comp Biol Med 1989; 19: 181-188.

A.M. Oudesluys-Murphy, W.C.J. Hop, C. de Groot. Umbilical cord separation in twins. Early Human Dev 1989; 19: 241-245.

J.B. Degener, M. Vogel, M.F. Michel, M. Mutsaers, W.C.J. Hop. The efficacy of the combination of teicoplanin or flucloxacillin with netilmicin in the treatment of Staphylococcus Aureus bacteremia. J Antimicrobial Chemoth 1989; 23: 899-904.

K.S.J.W. Ackaert, W.C.J. Hop, C.A. Heemskerk, F.H. Schröder. Risk factors in extracorporeal shockwave lithotripsy. Eur Urol 1989; 16: 349-353.

B. van den Berg, H. Stam, W.C.J. Hop. Effects of dietary protein content on weaning from the ventilator. Clinical Nutrition 1989; 8: 207-212.

I.A.L. Groenenberg, J.W. Wladimiroff, W.C.J. Hop. Fetal cardiac and peripheral arterial flow velocity waveforms in intrauterine growth retardation. Circulation 1989; 80: 1711-1717.

M.C.W. Scholtes, J.W. Wladimiroff, H. van Reijen, W.C.J. Hop. Uterine and ovarian flow velocity waveforms in the normal menstrual cycle; a transvaginal doppler study. Fertil Steril 1989; 6: 981-985.

B.C.J.M. Fauser, J.W. Bogers, W.C.J. Hop, F.H. de Jong. Bioactive and immunoreactive FSH in serum of normal and oligospermic men. Clin Endocrinol 1990; 32: 433-442.

D. Birnie, G.W.S. thoe Schwartzenberg, W.C.J. Hop, E.E.M. van Essen-Zandvliet, K.F. Kerrebijn. Does the outcome of the tidal breathing and dosimeter methods of assessing bronchial responsiveness in children with asthma depend on age? Thorax 1990; 45: 199-202.

J.H.W. de Wilt, J.N.M. IJzermans, W.C.J. Hop, J. Jeekel. De behandeling van recidief van hernia inguinalis. Ned T Geneesk 1990; 134: 531-534.

W. Baerts, W.P.F. Fetter, W.C.J. Hop, H.C.S. Wallenburg, R. Spritzer, P.J.J. Sauer. Cerebral lesions in preterm infants after tocolytic indomethacin. Developm Med Child Neurol 1990; 32: 910-918.

L. Dijkshoorn, A. van Ooyen, W.C.J. Hop, M. Theuns, M.F. Michel. Comparison of clinical Acinetobacter strains using a carbon source growth assay. Epidemiol Infect 1990; 104: 443-453.

A.M. Oudesluys-Murphy, J.C. den Hollander, W.C.J. Hop. Umbilical cord separation; histological findings and perinatal factors. Biol Neonate 1990; 58: 236-240.

A. Franken, F. Derkx, A. Man in't Veld, W. Hop, G. van Rens, E. Peperkamp, P. de Jong, M. Schalekamp. High plasma prorenin in diabetes mellitus and its correlation with some complications. J Clin Endocr Metab 1990; 71: 1008-1015.

M. Smit-Leijs, P. Kramer, R. Heijtink, W. Hop, S. Schalm. Hepatitis B vaccination of haemodialysis patients; randomized controlled trial comparing plasma-derived vaccine with and without pre-S2 antigen. Eur J Clin Invest 1990; 20: 540-545.

T. Pache, J. Wladimiroff, F. de Jong, W. Hop, B. Fauser. Growth patterns of non-dominant ovarian follicles during the normal menstrual cycle. Fertil Steril 1990; 54: 638-642.

T.E. Cohen-Overbeek, W.C.J. Hop, M. den Ouden, L. Pijpers, M.G.J. Jahoda, J.W. Wladimiroff. Spontaneous abortion rate and advanced maternal age; consequences for prenatal diagnosis. Lancet 1990; 336: 27-29.

G. Angelini, A. Fraser, M. Koning, J. Smyllie, W. Hop, G. Sutherland, P. Verdouw. Adverse hemodynamic effects and echocardiographic consequences of pericardial closure soon after sternotomy and pericardiotomy. Circulation 1990; 82 (suppl 4): 397-406.

H.W.J. van Marwijk, F. Bekker, W. Nolen, P. Jansen, J. van Nieuwkerk, W. Hop. Lithium augmentation in geriatric depression. J Affect Dis 1990; 20: 217-223.

F. de Waard-van der Spek, A. Oranje, S. Lillieborg, W. Hop, E. Stolz. Treatment of molluscum contagiosum using a lidocaine/prilocaine cream (EMLA) for analgesia. J Am Acad Dermatol 1990; 23: 685-688.

N.H.P.M. Jutte, W.C.J. Hop, R. Daane, C.E. Essed, W. Weimar, M.L. Simoons, E. Bos. Cytoimmunological monitoring of heart transplant recipients. Clin Transplantation 1990; 4: 297-300.

H. Haanen, H. Hoenderbos, L. van Romunde, W. Hop, C. Mallee, J. Terwiel. Controlled trial of hypnotherapy in the treatment of refractory fibromyalgia. J Rheumatol 1991; 18: 72-75.

W.C.J. Hop. Statgraphics bij het onderwijs in de Biostatistiek aan de Medische Faculteit in Rotterdam. Convex Courier 1991; 4: 3-6.

C. Jansen, B. Bonke, J. Klein, N. van Dasselaar, W. Hop. Failure to demonstrate unconscious perception during balanced anaesthesia by postoperative motor response. Acta Anaesth Scand 1991; 35: 407-410.

T.D. Pache, W.C.J. Hop, J.W. Wladimiroff, J. Schipper, B. Fauser. Transvaginal sonography and abnormal ovarian appearance in menstrual cycle disturbances. Ultrasound Med Biol 1991; 17: 589-593.

H. Balvert-Locht, J.W. Coebergh, W. Hop, H. Brolmann, M. Crommelin. Improved prognosis of ovarian cancer in the Netherlands during the period 1975-1985:a registry-based study. Gynecol Oncol 1991; 42: 3-8.

J. Rondeel, W. de Greef, W. Hop, T. Visser. Effect of cold exposure on the hypothalamic release of thyrotropin-releasing hormone and catecholamines. Neuroendocrinology 1991; 54: 477-481.

J. IJzermans, J. de Wilt, W. Hop, J. Jeekel. Recurrent inguinal hernia treated by classical hernioplasty. Arch Surg 1991; 126: 1097-1100.

I.A.L. Groenenberg, W.C.J. Hop, J.W. Wladimiroff. Doppler flow velocity waveforms in the fetal cardiac outflow tract; reproducibility of waveform recording and analysis. Ultrasound Med Biol 1991; 17: 583-587.

J. Coebergh, M. Verhagen-Teulings, M. Crommelin, L. v.d. Heijden, W. Hop. De overlevingskansen bij patienten met kanker gediagnostiseerd in 1975-1985in Zuidoost-Noord-Brabant en Noord-Limburg. Ned T Geneesk 1991; 21: 938-943.

B. Fauser, T. Pache, S. Lamberts, W. Hop, F. de Jong, K. Dahl. Serum bioactive and immunoreactive luteinizing hormone and follicle stimulating hormone levels in women with cycle abnormalities, with or without polycystic ovarian disease. J Clin Endocr Metab 1991; 73: 811-817.

P. van Broekhoven, W.C.J. Hop, E. Rasser, J.C. de Jongste, K.F. Kerrebijn. Comparison of FEV_1 and transcutaneous oxygen tension in the measurement of airway responsiveness to Methacholine. Pedr Pulm 1991; 11: 254-258.

I. Groenenberg, W. Baerts, W. Hop, J. Wladimiroff. Relationship between fetal cardiac and extra-cardiac Doppler flow velocity waveforms and neonatal outcome in intrauterine growth retardation. Early Human Dev 1991; 26: 185-192.

H. Hoogsteden, P. van Hal, J. Wijkhuys, W. Hop, A. Verkaik, C. Hilvering. Expression of the CD11/CD18 cell surface adhesion glycoprotein family on alveolar macrophages in smokers and nonsmokers. Chest 1991; 6: 1567-1571.

T. Pache, S. Chadha, L. Goorens, W. Hop, K. Jaarsma, H. Dommerholt, B. Fauser. Ovarian morphology in long term androgen treated female to male transsexuals. A human model for the study of polycystic ovarian syndrome. Histopathology 1991; 19: 445-452.

S.M. Hussain, M. Meradji, S. Robben, W.C.J. Hop. Plain film diagnosis in meconium plug syndrome, meconium ileus and neonatal Hirschprung's disease. Pediatr Radiol 1991; 21: 556-559.

J.E. Kist-van Holthe tot Echten, J. Nauta, W. Hop, M. de Jong, W. van Luijk, S. Ploos van Amstel, A. Roodhooft, C. Noordzij, E. Wolff. Protein intake can not be estimated from urinary urea excretion. Pediatr Nephrol 1992; 6: 85-87.

M. Meijssen, H. Tilanus, M. van Blankenstein, W. Hop, G. Ong. Achalasia complicated by oesophageal squamous cell carcinoma, a prospective study in 195 patients. Gut 1992; 33: 155-158.

Z.W. Sneller, W.C.J. Hop, P.J. Carpentier, F.H. Schröder. Prognosis and prostatic volume changes during endocrine management of prostate cancer patients:a longitudinal study. J Urol 1992; 147: 962-966.

A. de Jong, M. Terburg, W. Hart, J. Molenaar, P. Herbrink, W. Hop. Antilichamen tegen fosfolipiden bij jonge patienten met een cerebraal infarct. Revalidata 1992; 52A: 19-22.

P. Struyck, J. Wladimiroff, W. Hop, B. Simonazzi. Pulse pressure assessment in the human fetal descending aorta. Ultrasound Med Biol 1992; 1: 39-43.

M. Hoynck van Papendrecht, W. Hop, B. Langenhorst, F. Kothe, R. Marquet, J. Jeekel. A predeposit autologuous blood donation program in colorectal cancer patients. Vox Sang 1992; 62: 102-107.

M. Wilts, W. Hop, G. van der Heijden, K. Kerrebijn, J. de Jongste. Measurement of bronchial responsiveness in young children: comparison of transcutaneous oxygen tension and functional residual capacity during induced bronchoconstriction and -dilatation. Pediatr Pulm 1992; 12: 181-185.

T. Pache, W. Hop, F. de Jong, R. Leerentveld, H. van Geldrop, T. van Kamp, L. Gooren, B. Fauser. 17-Beta oestradiol, androstenedione and inhibin levels in fluid from individual follicles of normal and polycystic ovaries, and in ovaries from androgen treated female to male transsexuals. Clin Endocrinol 1992; 36: 565-571.

T. Pache, J. Wladimiroff, W. Hop, B. Fauser. How to discriminate between normal and polycystic ovaries: a transvaginal US study. Radiology 1992; 183: 421-423.

A. Franken, F. Derkx, P. Blankestijn, J. Janssen, C. Mannesse, W. Hop, F. Boomsma, R. Weber, E. Peperkamp, P. de Jong, M. Schalekamp. Plasma prorenin as an early marker of microvascular disease in patients with diabetes mellitus. Diabete Metabol 1992; 18: 137-143.

H.C. Hoogsteden, P.T.W. van Hal, J.M. Wijkhuis, W.C.J. Hop, C. Hilvering. Expression of the CD11/CD18 cell surface adhesion glycoprotein family and MCH Class-II antigen on blood monocytes and alveolar macrophages in interstitial lung diseases. Lung 1992; 170: 221-233.

D. Schoot, T. Pache, W. Hop, F. de Jong, B. Fauser. Growth patterns of ovarian follicles during induction of ovulation with decreasing doses of human menopausal gonadotropin following presumed selection in polycystic ovary syndrome. Fertil Steril 1992; 57: 1117-1120.

R. Singh, K. Mechelse, W. Hop, R. Braakman. Long term results of transplantation to repair median, ulnar and radial nerve lesions using a microsurgical interfascicular autogenous cable graft technique. Surg Neurol 1992; 37: 425-431.

F.C.E. van Knippenberg, J. Out, H. Tilanus, H. Mud, W. Hop, F. Verhage. Quality of life in patients with resected oesphageal cancer. Soc Sci Med 1992; 35: 139-145.

J. Blomjous, W. Hop, B. Langenhorst, F. ten Kate, W. Eykenboom, H. Tilanus. Adenocarcinoma of the gastric cardia: recurrence and survival after resection. Cancer 1992; 70: 569-574.

J.H. Ovelgönne, A.W.J.M. Bol, W.C.J. Hop, R. van Wijk. Mechanical agitation of very dilute antiserum against IgE has no effect on basophil staining properties. Experientia 1992; 48: 504-508.

H. Groen, J. Bogaard, A. Balk, S. Kho, W. Hop, C. Hilvering. Diffusion capacity in heart transplant recipients. Chest 1992; 102: 456-460.

J. Klinkenbijl, G. van der Schelling, W. Hop, R. van Pel, H. Bruining, J. Jeekel. The advantages of pylorus-preserving pancreatoduodenectomy in malignant disease of the pancreas and periampullary region. Ann Surg 1992; 216: 142-145.

F. de Jongh, H. Janssen, R. de Man, W. Hop, S. Schalm, M. van Blankenstein. Survival and prognostic indicators in hepatitis B surface antigen-positive cirrhosis of the liver. Gastroenterol 1992; 103: 1630-1635.

E.H. de Vries, A.Z. Ginai, S.G.F. Robben, W.C.J. Hop. Pelvic phleboliths, is there an association with diverticulitis? Br J Radiol 1992; 65: 868-870.

J.E. Kist-van Holthe tot Echten, J. Nauta, G. Boerma, W. Hop, C. Noordzij, E. Wolff. Protein restriction affects fat intake and serum lipids in children with chronic renal failure. Miner Electrol Metab 1992;18:207-211.

B. Fauser, T. Pache, W. Hop, F. de Jong, K. Dahl. The significance of a single serum LH measurement in women with cycle disturbances; discrepancies between immunoreactive and bioactive hormone estimates. Clin Endocrinol 1992;37:445-452.

M.B.E. Menke-Pluymers, N.W. Schoute, A.H. Mulder, W.C.J. Hop, M. van Blankenstein, H.W. Tilanus. Outcome of surgical treatment of adenocarcinoma in Barrett's oesophagus. Gut 1992;11:1454-1458.

K. van der Mooren, J.W. Wladimiroff, W.C.J Hop. Reproducibility of fetal cardiac flow velocity waveforms at atrioventricular level. Ultrasound Med Biol 1992;10:827-830.

J. Kist- van Holthe tot Echten, J. Nauta, W. Hop, M. de Jong, W. Reitsma, S. Ploos van Amstel, K. van Acker, C. Noordzij, E. Wolff. Protein restriction in chronic renal failure. Arch Dis Child 1993;68:371-375.

A.A.P.H. Verberne, W.C.J. Hop, A.B. Bos, K.F. Kerrebijn. Effect of a single dose of inhaled Salmeterol on baseline airway caliber and metacholine-induced airway obstruction in asthmatic children. J Allergy Clin Immunol 1993;91:127-134.

A.Z. Ginai, H.R. van Buuren, W.C.J. Hop, S.W. Schalm. Oesophageal varices, how reliable is a barium swallow? Brit J Radiol 1993;66:322-326.

B.E.M. van Essen-Zandvliet, W. Hop, H. de Jong, A. Ferwerda, K.F. Kerrebijn. Acute effect of an inhaled corticosteroid (Budesonide) on bronchial hyperresponsiveness to metacholine in children with asthma. Eur Respir J 1993;6:383-386.

A.J.M. van Deventer, H. van Vliet, L. Voogd, W. Hop, W. Goessens. Increased specificity of antibody detection in surgical patients with invasive candidiasis with cytoplasmic antigens depleted of mannan residues. J Clin Microbiol 1993;31:994-997.

A.W.L. de Jong, W. Hart, M. Terburg, J.L. Molenaar, P. Herbrink, W. Hop. Cardiolipin antibodies and lupus anticoagulant in young patients with a cerebrovascular accident in the past. Neth J Med 1993;42:93-98.

M.B.E. Menke-PLuymers, N. Schoute, A. Mulder, W. Hop, M. van Blankenstein, H. Tilanus. Overlevingsduur van patienten met een adenocarcinoom in een Barrett-oesofagus. Ned T Geneesk 1993;137:460-464.

I.A.L. Groenenberg, W.C.J. Hop, J.W. Boogers, J.G. Santema, J.W. Wladimiroff. The predictive value of Doppler flow velocity waveforms in the development of abnormal fetal heart rate traces in intrauterine growth retardation; a longitudinal study. Early Human Dev 1993;32:151-159.

T.D. Pache, F.H. de Jong, W.C.J. Hop, B.C.J.M. Fauser. Association between ovarian changes assessed by transvaginal sonography and clinical and endocrine signs of the polycystic ovary syndrome. Fertil Steril 1993;59:544-549.

O.R.C. Busch, W.C.J. Hop, M.A.W. Hoynck van Papendrecht, R.L. Marquet, J. Jeekel. Blood transfusions and prognosis in colorectal cancer. New Eng J Med 1993;328:1372-1376.

H.W. Tilanus, W.C.J. Hop, B.L.A.M. Langenhorst, J.J.B. van Lanschot. Esophagectomy with or without thoracotomy. Is there any difference? J Thor Cardiovasc Surg 1993;105:898-903.

H.C. Hoogsteden, P.T.W. van Hal, J.M. Wijkhuis, W.C.J. Hop, C. Hilvering. Differences in expression of monocyte/macrophage surface antigens in peripheral blood and bronchoalveolar lavage cells in interstitial lung diseases. Lung 1993;171:149-160.

F.C. van Biezen, P.A.G.M. Bakx, V.H. Devilleneuve, W.C.J. Hop. Scoliosis in children after thoracotomy for aortic coarctation. J Bone Joint Surg 1993;4:514-518.

H.C. Hoogsteden, M.A. Versnel, W. Hop, T.H. van der Kwast, C. Hilvering. Immunologic characterization of the leucocytes in pleural fluid in malignant mesothelioma. Eur Respir Rev 1993;3:30-32.

A.E. Bonebakker, B. Bonke, J. Klein, G. Wolters, W.C.J. Hop. Implicit memory during balanced anaesthesia; Lack of evidence. Anaesthesia 1993;48:657-660.

M.B.E. Menke-Pluymers, W.C.J. Hop, J. Dees, M. van Blankenstein, H.W. Tilanus. Risk factors for the development of an adenocarcinoma in columnar-lined (Barrett) esophagus. Cancer 1993;72:1155-1158.

O.R.C. Busch, R.L. Marquet, W.C.J. Hop, J. Jeekel. Blood transfusions and prognosis in colorectal cancer. New Eng J Med 1993; 329:1355-1356(letter).

D.C. Schoot, W.C.J. Hop, T.D. Pache, F.H. de Jong, B.C.J.M. Fauser. Growth of the dominant follicle is similar to normal in patients with gonadotrophin-stimulated polycystic ovary syndrome exhibiting monofollicular developement during a decremental dose regimen. Acta Endocrinol 1993;129:126-129.

L. Vandenbossche, W.C.J. Hop, J.C. de Jongste. Bronchial responsiveness to inhaled metabisulphite in asthmatic children increases with age. Pediatr Pulm 1993; 16: 236-242.

H.J.M. Groen, J.M. Bogaard, A.H.M.M. Balk, S.G. Kho, W.C.J. Hop, C. Hilvering. Diffusion capacity in heart transplant recipients. Chest 1993; 104: 1938-139.

European Multicentre Study Group. The results of a randomized double blind controlled trial evaluating malotilate in primary biliary cirrhosis. J Hepatology 1993; 17: 227-235.

J.W. Coebergh, M.A. Crommelin, L.H. van der Heijden, W.C.J. Hop, M.T. Verhagen-Teulings. Trends in cancer survival in the southeastern Netherlands since 1975; the Eindhoven Cancer Registry experience. Health Rep 1993; 5: 105-110.

J.W. Coebergh, A. van der Does, J.F. van Weerden, J.A. Rammeloo, E.R. van Wering, W.A. Kamps, W.C.J. Hop. Geen eenduidig verband tussen opleidingsniveau van ouders en de prognose van kinderen met acute leukaemie in Nederland 1973-79. In: Sociaal economische gezondheidsverschillen onderzocht (red: J.P. Mackenbach) 1993.

D.W. Paul, J.M. Bogaard, W.C.J.Hop. The bronchoconstrictor effect of strenuous exercise at low temperatures in normal athletes. Int J Sports Med 1993; 14: 433-436.

J.M.M. Gijtenbeek, W.C.J. Hop, R. Braakman, C.J.J. Avezaat. Surgery for intracranial meningiomas in elderly patients. Neurosurgery 1993; 4: 291-295.

P.F.J. Donderwinkel, D.C. Schoot, T.D. Pache, F.H. de Jong, W.C.J. Hop, B.C.J.M. Fauser BCJM. Ovary syndrome patients using exogeneous gonadotrophins in combination with a gonadotrophin-releasing hormone agonist. Human Reprod 1993; 8: 2027-2032.

O.R.C. Busch, W.C.J. Hop, R.L. Marquet, J. Jeekel. Prognostic impact of blood transfusions on disease-free survival in colorectal carcinoma. Scand J Gastroenterol 1993; 28 (suppl 200): 21-23.

A.J.M. van Deventer, H.J.A. van Vliet, W.C.J. Hop, W.H.F. Goessens. Diagnostic value of anti-candida enolase antibodies. J Clin Microbiol 1994; 32: 17-23.

M. Meradji, S.M. Hussain, S.G.F. Robben, W.C.J. Hop. Plain film diagnosis in intussusception. Br J Radiol 1994; 67: 147-149.

O.R.C. Busch, W.C.J. Hop, R.L. Marquet, J. Jeekel. Autologous blood and infections after colorectal surgery. Lancet 1994; 343: 668-669.

M.J. Noordam, R. Heydanus, W.C.J. Hop, F.M.E. Hoekstra, J.W. Wladimiroff. Doppler colour flow imaging of fetal intracerebral arteries and umbilical artery in the small for gestational age fetus. Br J Obstet Gynaecol 1994; 101: 504-508.

O.R.C. Busch, R.L. Marquet, W.C.J. Hop, J. Jeekel. Colorectal cancer recurrence and peri-operative blood transfusions; a critical reappraisal. Seminars Surg Oncol 1994; 10: 195-199.

J.N. van den Anker, W.C.J. Hop, R. de Groot, B.J. van der Heijden, H.M. Broerse, J. Lindemans, P.J.J. Sauer. Effects of prenatal exposure to Bethamethasone and Indomethacin on the glomerular filtration rate in the preterm infant. Pediatr Res 1994; 36: 578-581.

M.J. Noordam, F.M.E. Hoekstra, W.C.J. Hop, J.W. Wladimiroff. Doppler colour flow imagimg of fetal intracerebral arteries relative to fetal behavioural states in normal pregnancy. Early Human Dev 1994; 39: 49-56.

I.M. Risseeuw-Appel, I. Dekker, W.C.J. Hop, K. Hahlen. Minimal effects of E-coli and Erwinia Asparaginase on the coagulation system in childhood acute lymphoblastic leukemia: a randomized study. Med Pediatr Oncol 1994; 23: 335-343.

P. Sonneveld, W.C.J. Hop, A.H. Mulder, J.J. Michiels, G. Blijham, J. van der Lelie, J.W.M. Raemakers. K. Nieuwenhuis, C. van der Heul. Full-dose chemotherapy for non-Hodgkins's lymphoma in the elderly. Semin Hematol 1994; 2 (suppl 3): 9-12.

H.E. Lont, J.J. van Lanschot, W.C.J. Hop, W.M. Eijkenboom, P. Knegt, H.W. Tilanus. Reconstructie van de tractus digestivus met een vrij dunne-darm interonaat na totale larynx-farynxtirpatie, en follow-up resultaten. Ned T Geneesk 1994; 138: 1317-1321.

R.W. Luijendijk, C.C.A.P. Wauters, M.H.J. Voormolen, W.C.J. Hop, J. Jeekel. Intra-abdominale adhesies en vreemd-lichaam granulomen na eerdere laparotomie. Ned T Geneesk 1994; 138: 717-721.

H.W. Nab, W.C.J. Hop, M.A. Crommelin, H.M. Kluck, J.W.W. Coebergh. Improved prognosis of breast cancer since 1970 in south-eastern Netherlands. Br J Cancer 1994; 70: 285-288.

O.R.C. Busch, W.C.J. Hop, R.L. Marquet, J. Jeekel. Blood transfusions and local tumor recurrence in colorectal cancer; evidence for a noncausal relationship. Ann Surg 1994; 220: 791-797.

H. Jansen, W. Hop, A. van Tol, A.V.G. Bruschke, J.C. Birkenhager. Hepatic lipase and lipoprotein lipase are not major determinants of the low density lipoprotein subclass pattern in human subjects with coronary heart disease. Atherosclerosis 1994; 107: 45-54.

J.J.B. van Lanschot, W.C.J. Hop, H.H.J. Voormolen, R.A.J. van Deelen, J.G.A.M. Blomjous, H.W. Tilanus. Quality of palliation and possible benefit of extra-anatomic reconstruction in recurrent dysphagia and resection of carcinoma of the esophagus. J Am Coll Surg 1994; 179:705-713.

A. van der Gaast, C.H.H. Schoenmakers, T.C. Kok, B.G. Blijenberg, W.C.J. Hop, T.A.W. Splinter. Prognostic significance of tissue polypeptide-specific antigen (TPS) in patients with advanced non-small cell lung cancer. Eur J Cancer 1994; 30A: 12-.

M.B.E. Menke-Pluymers, A.H. Mulder, W.C.J. Hop, M. van Blankenstein, H.W. Tilanus. Dysplasia and aneuploidy as markers of malignant degeneration in Barrett's oesophagus. Gut 1994; 35: 1348-1351.

J.L.H.R. Bosch, W.C.J. Hop, A.Q.H.J. Niemer, C.H. Bangma, W.J. Kirkels, F.H. Schröder. Parameters of prostate volume and shape in a community based population of men 55 to 74 years old. J Urol 1994; 152: 1501-1505.

D. van den Ouden, P.J.T. Davidson, W.C.J. Hop, F.H. Schroeder. Radical prostatectomy as a monotherapy for locally advanced (stage T3) prostate cancer. J Urol 1994; 151: 646-651.

B. van den Berg, J.M. Bogaard, W.C.J. Hop. High fat, low carbohydrate, enteral feeding in patients weaning from the ventilator. Intensive Care Med 1994; 20: 470-475.

H.W. Nab, M.A. Crommelin, W.C.J. Hop, L.H. van der Heijden, J.W.W. Coebbergh. Changes in long term prognosis for breast cancer in a Dutch cancer registry. Br Med J 1994; 309: 803-806.

O.R.C. Busch, W.C.J. Hop, R.L. Marquet, J. Jeekel. Randomized clinical trial comparing the effect of allogeneic and autologous blood in colorectal cancer patients. In: Blood transfusion in cancer surgery. J.G.A. Houbiers (eds). Boerhaave Comm Postgrad Med Educat, Leiden 1994: 85-97.

J.N. van den Anker JN, W.C.J. Hop, R.C. Schoemaker, B.J. van der Heijden, H.J. Neijens, R. de Groot. Ceftazidime pharmacokinetics in preterm infants: effect of postnatal age and postnatal exposure to indometacin. Br J Clin Pharmacol 1995; 40: 439-443.

J.N. van den Anker, R.C. Schoemaker, W.C.J. Hop, B.J. van der Heijden, A. Weber, P.J.J. Sauer, H.J. Neijens. Ceftazidime pharmacokinetics in preterm infants; effects of renal function and gestational age. Clin Pharmacol Ther 1995; 58: 650-659.

C.H. Bangma, W.C.J. Hop, F.H. Schröder. Eliminating the need for per-operative frozen section analysis of pelvic lymph nodes during radiacal prostatectomy. Br J Urol 1995; 76: 595-599.

J.N. van den Anker, R. de Groot, H.M. Broerse, P.J.J. Sauer, B.J. van der Heijden, W.C.J. Hop, J. Lindemans. Assessment of glomerular filtration rate in preterm infants by serum creatinine: comparison with inulin clearance. Pediatrics 1995; 96: 1156-1158.

E.F.H. van Bommel, N.D. Bouvy, W.C.J. Hop, H.A. Bruining, W. Weimar. Use of APACHE II classification to evaluate outcome and response to therapy in acute renal failure in a surgical intensive care unit. Renal Failure 1995; 17: 731-742.

C.H. Bangma, W.C.J. Hop, F.H. Schröder. Serial prostate specific antigen measurements and progression in untreated confined (Stages T0 to 3NxM0, Grades 1 to 3) carcinoma of the prostate. J Urol 1995; 154: 1403-1406.

J.L.H.R. Bosch, W.C.J. Hop, W.J. Kirkels, F.H. Schröder. The international prostate symptom score in a community based sample of men between 55 and 74 years of age: prevalence and correlation of symptoms with age, prostate volume, flow rate and residual urine volume. Br J Urol 1995; 75: 622-630.

J.N. van den Anker, B.J. van der Heijden, W.C.J. Hop, R.C. Schoemaker, H.M. Broerse, H.J. Neijens, R. de Groot. The effect of asphyxia on the pharmacokinetics of ceftazidime in the term newborn. Pediatr Research 1995; 38: 808-811.

J.L.H.R. Bosch, W.C.J. Hop, W.J. Kirkels, F.H. Schröder. Natural history of benign prostatic hyperplasia: appropriate case definition and estimation of its prevalence in the community. Urology 1995; 46: 34-40.

O.R.C. Busch, W.C.J. Hop, R.L. Marquet, J. Jeekel. The effect of blood transfusions on survival after surgery for colorectal cancer. Eur J Cancer 1995; 31A: 1226-1228.

J.L.H.R. Bosch, W.C.J. Hop, C.H. Bangma, W.J. Kirkels, F.H. Schröder. Prostate Specific Antigen in a community based sample of men without prostate cancer: correlations with prostate volume, age, body mass index and symptoms of prostatism. Prostate 1995; 5: 241-249.

O.R.C. Busch, W.C.J. Hop, R.L. Marquet, J. Jeekel. Blood transfusions and local tumor recurrence in colorectal cancer - evidence of a non-causal relationship. Ann Surg 1995; 222: 757-758.

P.J.T. Davidson, W.C.J. Hop, K.H. Kurth, S.D. Fossa, H. Waehre, F.H. Schröder. Progression in untreated carcinoma of the prostate metastatic to regional lymph nodes (Stage T0 to 4, N1 to 3, M0, D1). J Urol 1995;154:2118-2122.

A. van der Gaast, J. Verweij, A.S.T. Planting, W.C.J. Hop, G. Stoter. Simple prognostic model to predict survival in patients with undifferentiated carcinoma of unknown primary site. J Clin Oncol 1995;13:1720-1725.

C.A.C. Hugen, A.M. Oudesluys-Murphy, W.C.J Hop. Serum sodium levels and probability of recurrent febrile convulsions. Eur J Pediatr 1995;154:403-405.

K.K. Krishnadath, H.W. Tilanus, H. van Blankenstein, W.C.J. Hop, R. Teijgeman, A.H. Mulder, F.T. Bosman, H. van Dekken. Accumulation of genetic abnormalities during neoplastic progression in Barrett's esophagus. Cancer Res 1995;55:1971-1976.

E.J.T. Luiten, W.C.J. Hop, J.F. Lange, H.A. Bruining. Controlled clinical trial of selective decontamination for the treatment of severe acute pancreatitis. Ann Surg 1995;222:57-65.

R.A. de Man, R.P. Sprey, H.G.M. Niesters, R.A. Heijtink, P.E. Zondervan, W. Hop, S.W. Schalm. Survival and complications in a cohort of patients with anti-Delta positive liver disease presenting in a tertiary referral clinic. J Hepatol 1995;23:662-667.

M.B.E. Menke-Pluymers, W.C.J. Hop, A.H. Mulder, H.W. Tilanus.
DNA ploidy as a prognostic factor for patients with an adenocarcinoma in Barrett's esophagus. Hepato Gastroenterology 1995;42:786-788.

O.R.C. Busch, W.C.J. Hop, R.L. Marquet, J. Jeekel. Autologous blood transfusions and prognosis in colorectal cancer surgery. J Clin Oncol 1995;13:1280.

H.W. Nab, H.M. Kluck, E.J.T. Rutgers, J.W. Coebergh, W.C.J. Hop. Long-term prognosis of breast cancer: an analysis of 462 patients in a general hospital in south east Netherlands. Eur J Surg Oncol 1995;21:42-46.

M.A. Noordzij, T.H. van der Kwast, G.J. van Steenbrugge, W.C.J Hop, F.H. Schröder. The prognostic influence of neuroendocrine cells in prostate cancer: results of a long-term follow-up study with patients treated by radical prostatectomy. Int J Cancer 1995;62:252-258.

A.L. Oen, M.F. de Boer, W.C.J. Hop, P. Knegt. Cervical metastasis from the unknown primary tumor. Eur Arch Otorhinolaryngol 1995;252:222-228.

E.J.P. van Santbrink, W.C.J. Hop, T.J.H.M. van Dessel, F.H. de Jong, B.C.J.M. Fauser. Decremental follicle-stimulating hormone and dominant follicle developement during the normal menstrual cycle. Fertil Steril 1995;64:37-43.

D.C. Schoot, W.C.J. Hop, F.H. de Jong, T.J.H.M. van Dessel, B.C.J.M. Fauser. Initial estradiol response predicts outcome of exogeneous gonadotropins using a step-down dose regimen for induction of ovulation in polycystic ovary syndrome. Fertil Steril 1995;64:1081-1087.

P. Sonneveld, M de Ridder, H. van der Lelie, K. Nieuwenhuis, A. Mulder, I. van Reijswoud, W. Hop, B. Lowenberg. Comparison of Doxorubicin and Mitroxantrone in the treatment of elderly patients with advanced diffuse Non-Hodgkin's lymphoma using CHOP versus CNOP chemotherapy. J Clin Oncol 1995;13:2530-2539.

H.A.W.M. Tiddens, P.D. Pare, J.C. Hogg, W.C.J. Hop, R. Lambertop, J.C. de Jongste. Cartilaginous airway dimensions and airflow obstruction in human lungs. Am J Respir Crit Care Med 1995;152:260-266.

D. Vroegindeweij, A.V. Tielbeek, J. Buth, F.P.G. Schol, W.C.J. Hop, G.H.M. Landman. Directional atherectomy versus balloon angioplasty in segmental femoropopliteal artery disease: two-year follow-up with color duplex scanning. J Vasc Surg 1995;21:255-269.

J.E. Kist-van Holthe tot Echten, J.G.M. Huijmans, W.C.J. Hop, L.A.H. Monnens, M.C.J.W. de Jong, C.M. Noordzij, R. Slotema, J. Nauta, E.D. Wolff. Intracellular amino acid concentrations in children with chronic renal insufficiency. Pediatr Nephrol 1996;10:46-50.

R.W. Luijendijk, D.C.D. de Lange, C.C.A.P. Wauters, W.C.J. Hop, J.J. Duron, J.L. Pailler, B.R. Camprodon, L. Holmdahl, H.J. van Geldorp, J. Jeekel. Foreign material in postoperative adhesions. Ann Surg 1996;3:242-248.

J.W.W. Coebergh, A. van der Does-van den Berg, W.C.J. Hop, F. van Weerden, J.A. Rammeloo, H.A. van Steensel, E.R. van Wering, W.A. Kamps. Small influence of parental educational level on the survival of children with leukaemia in The Netherlands between 1973 and 1979. Eur J Cancer 1996;2:286-289.

I.M. Risseeuw, I. Dekker, R.L. Stigter, W.C.J. Hop, K. Hählen. Progressive elevation of antithrombin III levels during acute lymphoblastic leukemia induction treatment, not followed by a decrease in antithrombin III after addition of asparaginase: arandomized study between Escherichia coli and Erwinia asparaginase. Int J Pediatr Hematol/Oncol 1996;3:69-74.

A.A.P.H. Verberne, W.C.J. Hop, F.B.M. Creyghton, R.W.G. van Rooij, M. van den Berg, J.C. de Jongste, K.F. Kerrebijn. Airway responsiveness after a single dose of salmeterol and during four months of treatment in children with asthma. J Allerg Clin Immunol 1996; 4: 938-946.

R.F. Kornelisse, J.A. Hazelzeth, H.F.J. Savelkoul, W.C.J. Hop, M.H. Suur, A.N.J. Borsboom, I.M. Risseeuw-Appel, E. van der Voort, R. de Groot. The relationship between plasminogen activator inhibitor-1 and proinflammatory and counterinflammatory mediators in children with meningococcal septic shock. J Infect Dis 1996; 173: 1148-1156.

A.Z. Ginai, F.C. van Biezen, P.A.M. Kint, H.Y. Oei, W.C.J. Hop. Digital subtraction arthography in preoperative evaluation of painful total hip arthroplasty. Skeletal Radiol 1996; 25: 357-363.

O.R.C. Busch, H.E. Lont, W.C.J. Hop, R.L. Marquet, J. Jeekel. The effect of blood donation on prognosis and infectious complications after colorectal cancer surgery. In: rhErythropoietin in cancer supportive treatment (J.F. Smyth et al, eds). Marcel Dekker, New York, 1996.

R.F. Kornelisse, K. Hoekman, J.J. Visser, W.C.J. Hop, J.G.M. Huijmans, P.J.C. van der Straaten, A.J. van der Heijden, R.N. Sukhai, H.J. Neijens, R. de Groot. The role of nitric oxide in bacterial meningitis in children. J Infect Dis 1996; 174: 120-126.

A. van der Burgh, J. Dees, W.C.J. Hop, M. van Blankenstein. Oesophageal cancer is an uncommon cause of death in patients with Barrett's oesophagus. Gut 1996; 39: 5-8.

C.C.M. Beerendonk, F.H.M. Derkx, A.P.M. Schellekens, W.C.J. Hop, P.A. van Dop. The influence of dietary sodium restriction on renal and ovarian renin and prorenin production during ovarian stimulation. Human Reprod 1996; 11: 956-961.

P. Honkoop, P.D. Siersema, H.W. Tilanus, L.P.S. Stassen, W.C.J. Hop, M. van Blankenstein. Benign anastomic strictures after transhiatal esophagectomy and cervical esophagogastrostomy: risk factors and management. J Thoracic Cardiovasc Surg 1996; 111: 1141-1147.