

# A novel nonparametric measure of explained variation for survival data with an easy graphical interpretation

## Ein neues nichtparametrisches Maß der Erklärten Variation für Überlebenszeitdaten mit einer einfachen grafischen Interpretation

### Abstract

**Introduction:** For survival data the coefficient of determination cannot be used to describe how good a model fits to the data. Therefore, several measures of explained variation for survival data have been proposed in recent years.

**Methods:** We analyse an existing measure of explained variation with regard to minimisation aspects and demonstrate that these are not fulfilled for the measure.

**Results:** In analogy to the least squares method from linear regression analysis we develop a novel measure for categorical covariates which is based only on the Kaplan-Meier estimator. Hence, the novel measure is a completely nonparametric measure with an easy graphical interpretation. For the novel measure different weighting possibilities are available and a statistical test of significance can be performed. Eventually, we apply the novel measure and further measures of explained variation to a dataset comprising persons with a histopathological papillary thyroid carcinoma.

**Conclusion:** We propose a novel measure of explained variation with a comprehensible derivation as well as a graphical interpretation, which may be used in further analyses with survival data.

**Keywords:** explained variation, survival data, novel measure, nonparametric, test of significance

### Zusammenfassung

**Einleitung:** Das Bestimmtheitsmaß kann bei Überlebenszeitdaten nicht verwendet werden um mithilfe einer Maßzahl anzugeben, wie gut ein Modell zu den vorliegenden Daten passt. Daher wurden in den letzten Jahren mehrere Maße der Erklärten Variation für Überlebenszeitdaten vorgeschlagen.

**Methoden:** Wir analysieren eines dieser Maße der Erklärten Variation bezüglich gewisser Minimierungseigenschaften und zeigen, dass diese für das Maß nicht erfüllt sind.

**Ergebnisse:** In Analogie zu der Kleinste-Quadrate-Methode aus der linearen Regression entwickeln wir ein neues Maß für kategorielle Kovariaten, welches nur auf dem Kaplan-Meier-Schätzer basiert. Dadurch ist das neue Maß komplett nichtparametrisch und besitzt eine einfache grafische Interpretation. Für das neue Maß stehen verschiedene Gewichtungsmöglichkeiten zur Verfügung und ein statistischer Signifikanztest kann angewendet werden. Abschließend bestimmen wir das neue Maß sowie weitere Maße der Erklärten Variation für die Personen eines Datensatzes mit einem histopathologisch gesicherten papillären Schilddrüsenkarzinom.

**Schlussfolgerung:** Wir schlagen ein neues Maß der Erklärten Variation mit einer eingängigen Herleitung sowie einer grafischen Interpretation vor, welches bei künftigen Auswertungen von Überlebenszeitdaten verwendet werden könnte.

Verena Weiß<sup>1</sup>  
Matthias Schmidt<sup>2</sup>  
Martin Hellmich<sup>1</sup>

1 Institute of Medical Statistics, Informatics and Epidemiology, University of Cologne, Germany

2 Department of Nuclear Medicine, University of Cologne, Germany

**Schlüsselwörter:** Erklärte Variation, Überlebenszeitdaten, neues Maß, nichtparametrisch, Signifikanztest

## 1 Introduction

In linear regression analysis one often makes use of the coefficient of determination to describe how good a model fits to the data. The coefficient of determination is a measure with values between 0 and 1 and is defined as the square of the multiple correlation coefficient [1]. Due to censoring and a potential skewness of the data one cannot use this measure for survival data. Therefore, several measures of explained variation for survival data have been proposed in recent years, most of which are not easy to grasp or to interpret. In this work we present a novel measure of explained variation with a comprehensible derivation as well as a graphical interpretation. Furthermore we construct the novel measure in a completely nonparametric way and propose the application of a test of significance.

One of the measures of explained variation for survival data is the measure  $V_1$  by Schemper [2]. This measure has, together with the measure  $V_2$  by Schemper [2], been applied in several clinical, epidemiological and biological applications [3], [4], [5], [6]. In the measure  $V_1$  the impact of covariates is determined by the comparison of the distances of a singleton survival curve, i.e. the survival curve of only one person, to the Kaplan-Meier estimator of the entire group of persons and the distances of a singleton survival curve to the survival curve which results from a Cox proportional hazard model with given covariates. We analyse for Schemper's measure  $V_1$  whether the distances employed in this measure are appropriate regarding minimisation aspects, which are chosen in analogy to the least squares method from linear regression analysis and the minimisation property of the median. In other words, we investigate whether  $V_1$  is based on a proper scoring rule [7], that again is based on a type of absolute error loss [8].

Based on the measure  $V_1$  by Schemper and the consideration of these minimisation aspects we derive a novel measure of explained variation  $V_1^*$  for categorical covariates with an easy graphical interpretation. On the one hand the novel measure defines the distance of a singleton survival curve to the Kaplan-Meier estimator in a different way and on the other hand it accounts for categorical covariates by using the distances of a singleton survival curve to the Kaplan-Meier estimator of the group to which the person belongs due to the covariates. The novel measure is based only on the Kaplan-Meier estimator and, as a consequence, is a completely nonparametric measure which can be depicted in a graphical way. Furthermore, different possibilities to weight the distances between a singleton survival curve and the Kaplan-Meier estimator at the event times are given and a statistical test of significance can be performed. This enables the comparison of the explained variation according to the novel measure  $V_1^*$  for two cat-

egorical covariates by inferential means. Finally, we apply the novel measure as well as further measures of explained variation to a dataset comprising 508 persons with differentiated papillary thyroid carcinoma.

## 2 Methods

### 2.1 The measure $V_1$ by Schemper

The survival process of person  $i$  with  $i = 1, \dots, n$  at time  $t_j$  with  $1 \leq j \leq k_i$  is given by  $S_i(t_j)$  and assumes a value of 1 if person  $i$  is at risk for the event at time  $t_j$ , of 0.5 if  $t_j$  is the exact failure time of person  $i$  and of 0 if person  $i$  has experienced the event before time  $t_j$ . If person  $i$  experiences the event, the term  $k_i$  is the total number of events in the sample. If person  $i$  is censored,  $k_i$  is the number of events prior to the censoring of person  $i$ . The measure  $V_1$  by Schemper [2] is defined as

$$V_1 = 1 - \frac{\sum_{i=1}^n k_i^{-1} \sum_j |S_i(t_j) - \hat{S}_i(t_j|Z_i)|}{\sum_{i=1}^n k_i^{-1} \sum_j |S_i(t_j) - \hat{S}_{KM}(t_j)|},$$

where  $\hat{S}_{KM}(t_j)$  denotes the Kaplan-Meier estimator at time  $t_j$  and  $\hat{S}_i(t_j|Z_i)$  the survival curve derived from a Cox model at time  $t_j$  with covariates  $Z_i$  of person  $i$ . The measure  $V_1$  compares the distances between a singleton survival curve and the value of the Kaplan-Meier estimator at the event times  $t_j$  (in the "denominator") with the distances between a singleton survival curve and the value of the survival curve derived from a Cox model with given covariates (in the "numerator") at the event times  $t_j$ . The definition of the measure  $V_1$  by Schemper and Stare [9] is identical to the definition above with the exception that the survival process  $S_i(t_j)$  assumes a value of 1 if person  $i$  is at risk for the event at time  $t_j$  and of 0 if person  $i$  has experienced the event before or experiences the event exactly at time  $t_j$ . We will use this definition by Schemper and Stare throughout this paper.

### 2.2 Examination of minimisation aspects of the measure $V_1$ by Schemper

To assess the relevance of the covariates the measure  $V_1$  by Schemper compares the distances between a singleton survival curve, i.e. the survival curve of only one person, and the value of the Kaplan-Meier estimator of the entire group of persons at the event times  $t_j$  (in the denominator) with the distances between a singleton survival curve and the value of the survival curve derived from a Cox model with given covariates (in the numerator) at the event times  $t_j$ . The distances to the value of the Kaplan-Meier estimator of the entire group of persons therefore serve as a reference for the relevance of the covariates. It would be desirable if in the denominator the vertical, absolute distances between the singleton

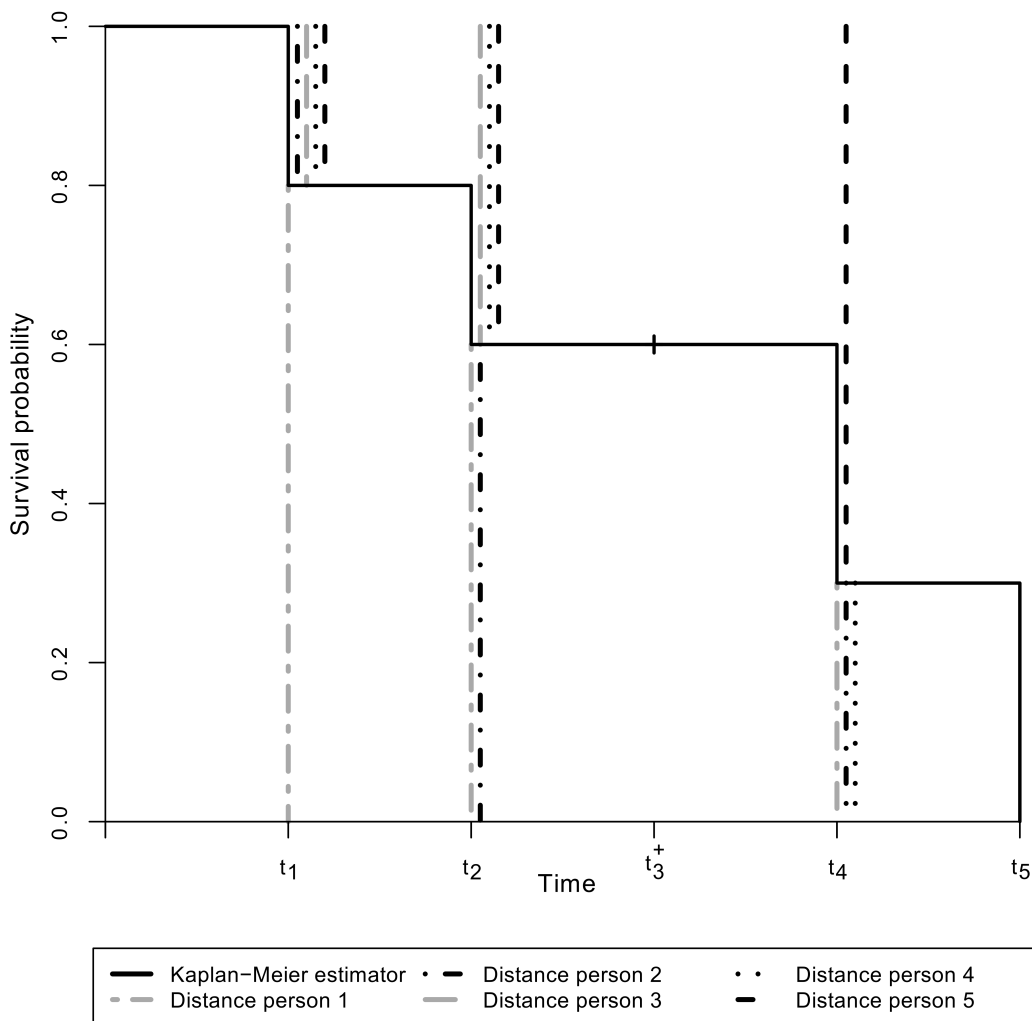


Figure 1: Distances in the measure  $V_1$  for the five persons in the example

survival curves and the value of the Kaplan-Meier estimator yield the Kaplan-Meier estimator as the optimal (minimising) step function, because the Kaplan-Meier estimator is a consistent estimator of the survival function with negligible bias [10]. Therefore it is desirable that the vertical, absolute distances generated in the denominator of  $V_1$  yield the Kaplan-Meier estimator as the optimal (minimising) step function. This is the case if at each event time the vertical, absolute distances above (“upper distances”) and below (“lower distances”) the Kaplan-Meier step function balance each other out, i.e. the sum of the distances above and the sum of the distances below the Kaplan-Meier step function have the same value at each event time. This approach arises from the minimisation property of the median, which implies that the median minimises the sum of the absolute differences from the single values to the median [11]. To investigate this circumstance we set up a theoretical example with five persons: person 1 experiences the event at time  $t_1$ , person 2 at time  $t_2$ , person 3 is censored at time  $t_3^+$ , person 4 experiences the event at time  $t_4$  and

person 5 at time  $t_5$ . The vertical, absolute distances defined in the denominator of the measure  $V_1$  for the five persons are displayed in Figure 1. One can see that the upper distances for persons at risk at an event time reach from the value 1 to the value of the Kaplan-Meier estimator at the event time. Furthermore if a person experiences the event a lower distance at the event time itself as well as at the subsequent event times is given and reaches from the value 0 to the value of the Kaplan-Meier estimator at the respective event time. At the top of Table 1 the absolute values of these distances above and below the Kaplan-Meier step function are given for each person at every event time. At time  $t_5$  person 5 experiences the event and no person is at risk for the event. For this reason no upper distances are generated at this event time. The upper respectively lower distances are accumulated over the five persons at each event time. It can be seen that these accumulated upper and lower distances do balance each other out at the event times  $t_1$  and  $t_2$  in contrast to the event time  $t_4$ . Therefore it can be stated that the distances in the measure  $V_1$  do not yield the

**Table 1: Distances above ( $U$ ) and below ( $L$ ) the Kaplan-Meier estimator in the measures  $V_1$  and  $V_1^*$  in the example**

		Distances measure $V_1$							
		Event time							
		$t_1$		$t_2$		$t_4$		$t_5$	
Person $i$		$U_i$	$L_i$	$U_i$	$L_i$	$U_i$	$L_i$	$U_i$	$L_i$
1		-	0.8	-	0.6	-	0.3	-	0
2		0.2	-	-	0.6	-	0.3	-	0
3		0.2	-	0.4	-	-	-	-	-
4		0.2	-	0.4	-	-	0.3	-	0
5		0.2	-	0.4	-	0.7	-	-	0
Sum		0.8	0.8	1.2	1.2	0.7	0.9	-	0
		Distances measure $V_1^*$							
		Event time							
		$t_1$		$t_2$		$t_4$		$t_5$	
Person $i$		$U_i$	$L_i$	$U_i$	$L_i$	$U_i$	$L_i$	$U_i$	$L_i$
1		-	0.8	-	-	-	-	-	-
2		0.2	-	-	0.6	-	-	-	-
3		0.2	-	0.2	-	-	-	-	-
4		0.2	-	0.2	-	-	0.3	-	-
5		0.2	-	0.2	-	0.3	-	-	0
Sum		0.8	0.8	0.6	0.6	0.3	0.3	-	0

Kaplan-Meier estimator as the optimal (minimising) step function at every event time, because the upper and lower distances do not balance each other out at every event time.

### 3 Results

#### 3.1 Definition of a novel, nonparametric measure of explained variation $V_1^*$

In this section a novel measure of explained variation  $V_1^*$  is developed. In this measure the vertical, absolute distances are defined in such a way, that they balance each other out at every event time. In the denominator of the novel measure  $V_1^*$  an upper distance at an event time  $t_j$  for person  $i$  is defined as the vertical, absolute distance between the value of the Kaplan-Meier estimator at the preceding event time  $t_{j-1}$  and the value of the Kaplan-Meier estimator at the event time  $t_j$  itself. In contrast to this each upper distance in the measure  $V_1$  starts at the value of the singleton survival curve of a person, i.e. at the value 1. Furthermore in the denominator of the novel measure  $V_1^*$  a lower distance is determined only at the event time  $t_j$  of a person and reaches from 0 to the value of the Kaplan-Meier estimator at this event time. At the subsequent event times no lower distances are considered for the person in contrast to the measure  $V_1$ . The upper and the lower distances in the novel measure  $V_1^*$  are analogous to the definition of the Kaplan-Meier estimator, in which the value of the estimation at an event

time is dependent on the value of the estimation at the preceding event time and a person who experiences the event will not be considered in the estimation after the event time. Reconsidering the example with five persons, one can see that the vertical, absolute distances as defined in the novel measure  $V_1^*$  yield the Kaplan-Meier estimator as the optimal (minimising) step function at every event time. For this purpose the vertical, absolute distances defined in the measure  $V_1^*$  are illustrated for the five persons in Figure 2. For example, for the event time  $t_2$  it can be seen that the upper distances of the persons 3, 4 and 5, who are at risk for the event at this time, reach from the value of the Kaplan-Meier estimator at the preceding event time  $t_1$  to the value of the Kaplan-Meier estimator at time  $t_2$ . Furthermore for persons experiencing the event, the lower distance contributes at the event time only. At subsequent event times no further lower distances are considered for these persons. At the bottom of Table 1 the values of the distances above and below the Kaplan-Meier estimator for each person at each event time are given and accumulated over the persons at each event time. In contrast to the measure  $V_1$  the accumulated upper and lower distances in the measure  $V_1^*$  do balance each other out at every event time and, as a consequence, yield the Kaplan-Meier estimator as the optimal (minimising) step function.

We take a further modification of the measure  $V_1$  concerning the consideration of a categorical covariate  $Z_1$  into account. In the numerator of the measure  $V_1$  by Schemper the impact of a covariate is assessed by the vertical, absolute distances of the singleton survival curve of person  $i$  to the survival curve yielded by a Cox model with given

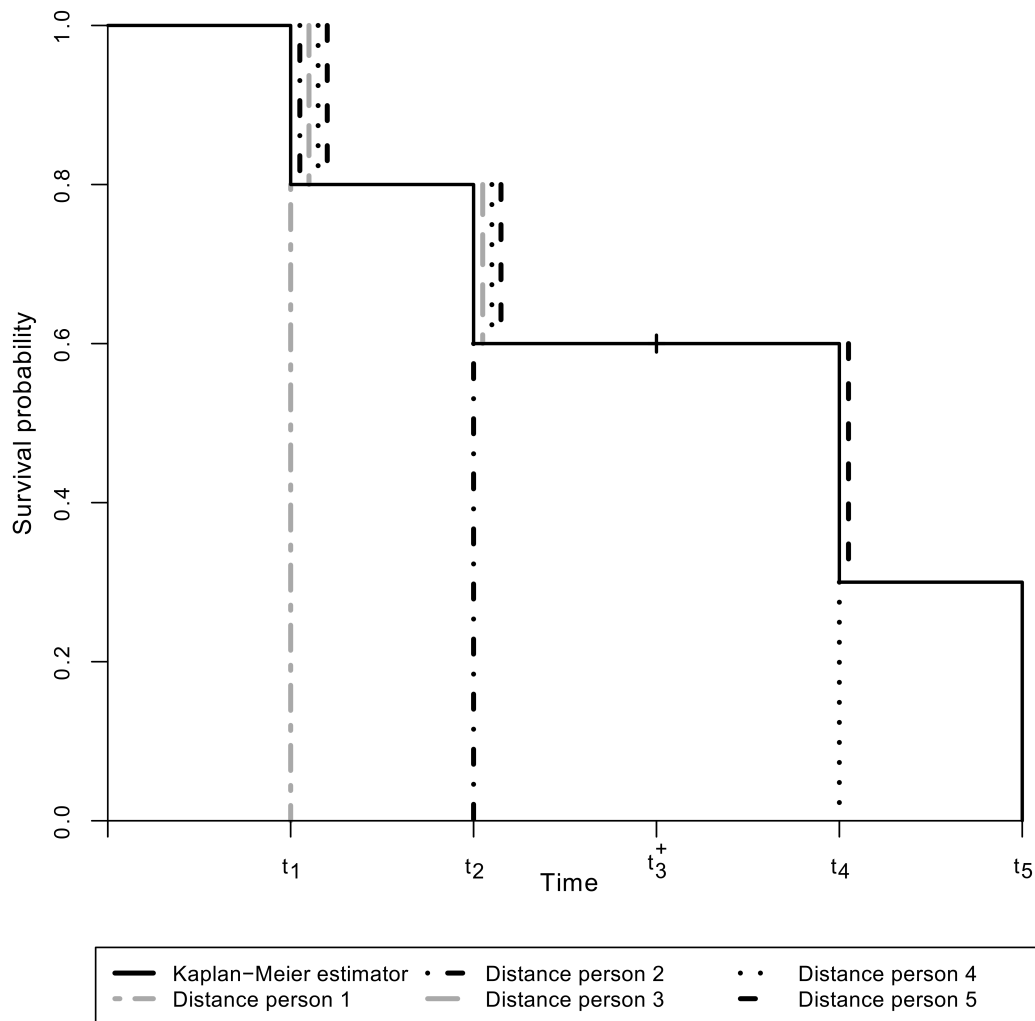


Figure 2: Distances in the measure  $V_1^*$  for the five persons in the example

a covariate of the person. As a consequence the proportional hazards assumption must be satisfied for the measure  $V_1$ . We modify the measure  $V_1$  by assessing the distances between the singleton survival curve of a person  $i$  to the value of the Kaplan-Meier estimator of the group, to which person  $i$  belongs according to the covariate. Therefore, the novel measure  $V_1^*$  is a completely nonparametric measure, because it is based only on the Kaplan-Meier estimator.

The novel measure  $V_1^*$  for a categorical covariate  $Z_1$  is defined as

$$V_1^*(Z_{1,i}) = 1 - \frac{\sum_{i=1}^n \sum_j w_j |\min[S_i(t_j), \hat{S}_{KM_{Z_{1i}}}(t_{j-1})] - \hat{S}_{KM_{Z_{1i}}}(t_j)|}{\sum_{i=1}^n \sum_j w_j |\min[S_i(t_j), \hat{S}_{KM}(t_{j-1})] - \hat{S}_{KM}(t_j)|}$$

In this formula the term  $S_i(t)$  denotes the survival process of person  $i$  at time  $t$ ,  $\hat{S}_{KM}(t)$  the Kaplan-Meier estimator of the entire group at time  $t$  and  $\hat{S}_{KM_{Z_{1i}}}(t_j)$  denotes the Kaplan-Meier estimator of the group to which person  $i$  belongs according to the covariate at time  $t_j$ .

Due to the terms containing the minimum it is ensured that the upper respectively the lower distances to the Kaplan-Meier estimator of the entire group respectively of the group to which the person belongs according to the covariate  $Z_1$  at time  $t_j$  are provided. An upper distance thereby reaches from the value of the Kaplan-Meier estimator at the preceding event time  $t_{j-1}$  to the value of the Kaplan-Meier estimator at the event time  $t_j$  and a lower distance from 0 to the value of the Kaplan-Meier estimator at the event time  $t_j$ . For a person who experiences the event the index  $j$  contains all event times up to and including the event time of the person. For a censored person the index  $j$  contains all event times up to the censoring time of the person (including this time if another person experiences the event at this time). Furthermore, different weighting possibilities  $w_j$  are available, which will be specified in detail in the next section.



### 3.2 Different weighting possibilities

The available weighting possibilities for the novel measure  $V_1^*$  are analogous to the weighting possibilities of the Logrank test. It must be ensured that, despite the weighting term, the upper and lower distances balance each other out at every event time. The following weighting terms  $w_j$  fulfil this requirement:

$w_j = 1$ : The vertical, absolute distances are weighted equally with the value 1 at every event time, which corresponds to the weighting in the Logrank test itself [12].

$w_j = \sqrt{Y_j}$ : The vertical, absolute distances are weighted with  $w_j = \sqrt{Y_j}$ , i.e. with the square root of the number of persons at risk for the event at the time  $t_j$ . This weighting possibility corresponds to the weighting in the Tarone-Ware test [13].

$w_j = Y_j$ : The vertical, absolute distances are weighted with  $Y_j$ , i.e. with the number of persons at risk for the event at the time  $t_j$ . This weighting possibility corresponds to the weighting in the Breslow test [12].

In the last two weighting possibilities the vertical, absolute distances at the beginning of the study, i.e. at times with a larger number of persons at risk, are more emphasized than the vertical, absolute distances at the end of the study, i.e. at times with fewer persons at risk.

### 3.3 Test of significance

In this section we suggest how the explained variation of a covariate  $Z_1$  can be compared to the explained variation of a covariate  $Z_2$  by inferential means. Assuming that the two covariates are gathered from the same persons, the test of significance has to be a test for paired samples. In the previous section we noted that the novel measure of explained variation  $V_1^*$  is a completely nonparametric measure and consequently the corresponding test of significance should be nonparametric as well. Therefore we apply the Wilcoxon test for paired samples [14].

As can be detected in the formula, the covariates are accounted for solely in the numerator of the measure  $V_1^*$ . For the covariates  $Z_1$  and  $Z_2$  the corresponding terms in the numerator, i.e. the accumulated weighted sum of distances of a person  $i$  according to the covariate  $Z_1$  and  $Z_2$  respectively, are

$$D^*(Y_i | Z_{i1}) = \sum_j w_j \min [S_i(t_j), \hat{S}_{KM_{Z_{i1}}}(t_{j-1})] - \hat{S}_{KM_{Z_{i1}}}(t_j)$$

and

$$D^*(Y_i | Z_{i2}) = \sum_j w_j \min [S_i(t_j), \hat{S}_{KM_{Z_{i2}}}(t_{j-1})] - \hat{S}_{KM_{Z_{i2}}}(t_j).$$

To compare two covariates due to their extent of explained variation, a Wilcoxon test for paired samples is performed over the differences of these accumulated distances, i.e. over the terms  $D_i^* = D^*(Y_i | Z_2) - D^*(Y_i | Z_1)$ .

### 3.4 Medical application

We apply the novel measure  $V_1^*$  as well as other measures of explained variation to a dataset comprising 508 per-

sons with a histopathological papillary thyroid carcinoma. These persons were treated by total thyroidectomy and subsequent radioiodine therapy with 3.7 GBq between Januar 1990 and June 2005 at the Department of Nuclear Medicine, University Hospital of Cologne in Germany. The tumour of each person was retrospectively classified according to the fifth [15], the sixth [16] and the seventh version [17] of the UICC tumour classification system. The UICC classification system describes the anatomical extent of a malignant disease and is based on the assessment of three components. The first component "T" describes the extent of the primary tumour, the second component "N" the absence or presence and extent of regional lymph node metastases and the third component "M" the absence or presence of distant metastases. Between the three versions of the UICC tumour classification system the margins, which assign a thyroid carcinoma to the respective tumour category of component "T", were redefined. More precisely the margins for T1 tumours were extended from 1 centimetre in the fifth version to 2 centimetres in the sixth version. Furthermore a minimal extrathyroidal growth, which defines a T4 tumour in the fifth version, defines a T3 tumour in the sixth version. In the seventh version the margins of the sixth version were not changed, but T1 tumours are subdivided into the categories a and b, for tumour diameters  $\leq 1$  centimetre respectively  $> 1$  centimetre. As a consequence the fifth and sixth version each comprise four categories, whereas the seventh version consists of five categories. Based on this dataset we seek to answer the question whether the fifth, the sixth or the seventh version of the UICC classification system explains the most variation for the outcome variable "time to the occurrence of a distant metastasis". In the 508 persons with a papillary thyroid carcinoma 25 distant metastases were diagnosed. A detailed description of the dataset, which comprises as a whole 636 persons with either a papillary or a follicular thyroid carcinoma, can be found in Meixner et al. [18]. Since only routinely collected data were retrospectively evaluated ethical review was not required.

The allocation of the tumours of the 508 persons with a papillary thyroid carcinoma to the categories according to the fifth and the sixth version of the UICC tumour classification system is given in Table 2. One can see, that the tumour of 227 persons is assigned to different categories by the fifth respectively the sixth version. The seventh version subdivides the tumours allocated to the category pT1 by the sixth version merely into the categories pT1a (133 tumours) and pT1b (127 tumours). The 25 events occur as follows in the categories pT1 (respectively pT1a and pT1b), pT2, pT3 and pT4: 2, 6, 3, 14 events (fifth version), 4, 4, 8, 9 events (sixth version) and 2, 2, 4, 8, 9 events (seventh version). To analyse whether the fifth or the sixth version explains more variation, the novel measure  $V_1^*$  with weighting term  $w_j=1$  is calculated on the one hand with the fifth version as a covariate and on the other hand with the sixth version as a covariate. For the time to the occurrence of a distant metastasis the measure  $V_1^*$  has a value of 0.0239 for the fifth version

**Table 2: Allocation of the tumours to the categories of the fifth and sixth version of the UICC tumour classification system**

		UICC-6				
		pT1	pT2	pT3	pT4	Sum
UICC-5	pT1	133	0	0	0	133
	pT2	127	101	0	0	228
	pT3	0	0	23	0	23
	pT4	0	0	100	24	124
	Sum	260	101	123	24	508

**Table 3: Values of various measures of explained variation considering one of three versions of the UICC tumour classification system as a covariate for the outcome variable**

Covariate	Measure of explained variation							
	$V_1^*$	$V_1$	$V_2$	$V$	$R_{LR}^2$	$\bar{R}_{LR}^{-2}$	$\rho_W^2$	$\rho_{W,A}^2$
UICC-5	0.0239	0.0298	0.0045	0.0338	0.0342	0.0752	0.3331	0.3745
UICC-6	0.0695	0.0768	0.0059	0.0870	0.0609	0.1341	0.4532	0.5190
UICC-7	0.0700	0.0768	0.0059	0.0869	0.0609	0.1341	0.5322	0.5735
Weighting term for $V_1^*$ : $w_j=1$								

and of 0.0695 for the sixth version of the UICC tumour classification system. The corresponding test of significance, i.e. the Wilcoxon test for paired samples, has a p-value less than 0.0001. In summary the explained variation has a larger value according to the sixth version than the explained variation according to the fifth version for the outcome variable and the difference is statistically significant at the significance level  $\alpha=0.05$ . Consequently it can be stated that the sixth version explains significantly more variation than the fifth version of the UICC tumour classification system for the outcome variable. For the seventh version the measure  $V_1^*$  has a value of 0.0700 and explains in comparison to the sixth version significantly more variation ( $p<0.0001$ ). To put the results of the measure  $V_1^*$  in a broader context the values of further measures of explained variation are considered, these are the measures  $V_1$  and  $V_2$  by Schemper [2], the measure  $V$  by Schemper and Henderson [19], a measure  $R_{LR}^2$  based on the likelihood ratio test [20], the measure  $\bar{R}_{LR}^{-2}$  by Nagelkerke [21] and the measure  $\rho_W^2$  by Kent and O'Quigley [22] as well as the approximation thereof  $\rho_{W,A}^2$  [22]. To compute the measures of explained variation the statistic software packages R 2.15.1 (measures  $V_1^*$ ,  $\bar{R}_{LR}^{-2}$ , SAS® 9.3 (TS1M0) with SAS/STAT 9.3 (measures  $V_1$ ,  $V_2$ ,  $R_{LR}^2$ ,  $\rho_W^2$ ,  $\rho_{W,A}^2$ ) and SAS® 9.4 (TS1M1) with SAS/STAT 13.1 (measure  $V$ ) are used [23], [24], [25]. Details concerning the computation of the measures can be found in the Appendix.

The values of the measures of explained variation for the outcome variable with the fifth, the sixth and the seventh version of the UICC tumour classification system respectively, as a covariate are given in Table 3. The values of the measure  $V_1^*$  and the measure  $V_1$  are of a quite similar dimension, as well as the values of the measure  $V_2$ , which are just slightly smaller than the values of the measures

$V_1^*$  and  $V_1$ . Contrarily, the values of the measures  $V$  and  $R_{LR}^2$  are predominantly larger, followed by the values of the measure  $\bar{R}_{LR}^{-2}$ . The values of the measures by Kent and O'Quigley  $\rho_W^2$  and  $\rho_{W,A}^2$  are considerably larger than the values of all other measures of explained variation. Furthermore one can see that for all measures of explained variation the values according to the sixth version are larger than the values according to the fifth version. The values according to the seventh version are, in comparison to the values of the sixth version, identical or slightly larger.

## 4 Discussion

The novel measure of explained variation  $V_1^*$  is a comprehensible measure which can be depicted in a graphical way. As a reference for the relevance of a categorical covariate the vertical, absolute distances between the singleton survival curves and the value of the Kaplan-Meier estimator are defined in this measure in such a way that the distances above and below the Kaplan-Meier estimator balance each other out at every event time. This property is desirable in analogy to the least squares method from linear regression analysis and the minimisation property of the median. Furthermore in the measure  $V_1^*$  the explained variation is determined by the comparison of these distances to the Kaplan-Meier estimator of the group to which the person according to the covariate belongs. Thereby the novel measure is based only on the Kaplan-Meier estimator and, as a consequence, is a completely nonparametric measure. The sole assumption is the validity of the "independent censoring assumption", which requires that the censored persons at a time are representative for the persons at risk at this time [26]. The interpretation of the distances

applied in the measure  $V_1^*$  arises from the analogy to the definition of the Kaplan-Meier estimator, in which the value of the estimation at an event time is dependent on the value of the estimation at the preceding event time and a person who experiences the event will not be considered in the estimation after the event time. Furthermore, different weighting possibilities are available for the measure  $V_1^*$ . These are comparable to the weighting possibilities of the Logrank test. One can therewith weight the distances at each event time equally or one can weight the distances at the early event times with many persons at risk more strongly than the event times at the end of the study with fewer persons at risk.

A further aspect of this work is the application of the Wilcoxon test for paired samples to compare the explained variation of two covariates. In addition to the mere comparison of the values of the measure  $V_1^*$  it can be analysed whether one covariate explains significantly more variation than the other covariate, i.e. whether the difference is statistically significant or not. This significance test could be implemented for the measure  $V_1$  as well. However, for both measures the assumptions required for the Wilcoxon test for paired samples are not entirely met (i.e. the independence of paired differences may be violated due to reference to a common Kaplan-Meier curve), respectively need to be assumed (symmetric continuous distribution of the paired differences around the median 0). The latter assumption is not required for the exact binomial sign test, while both requirements can be waived for an appropriate permutation test.

Further limitations of the measure  $V_1^*$  must be stated. Continuous variables have to be categorised, which entails a loss of information, before being incorporated in the measure  $V_1^*$ . Note that an upper limit for the number of categories does not exist, i.e. the measure can always be calculated. However in order to get valid and reproducible results the “10 events per variable/category rule” may be applied [27]. In the extreme case where every person is in his or her own category the explained variation would amount to 1. Furthermore the quantity which is supposed to be estimated by the novel measure, i.e. the estimand [28], is hard to grasp. However the novel measure  $V_1^*$  has the following advantages: It is a proper scoring rule, which yields the Kaplan-Meier estimator as the optimal minimising step function at every event time. Afterwards the measure combines the respective distances over all event times with the consequence that it does not solely correspond to a single event time. Furthermore the measure has a comprehensible derivation with a graphical interpretation and is a completely nonparametric measure. Moreover the covariates in the medical application, which is described in this manuscript, are categorical. The possibility to incorporate continuous covariates into the measure is not needed for this medical application.

In contrast to the measure  $V_1^*$ , the distances above and below the Kaplan-Meier estimator in the measure  $V_1$  do not balance each other out at every event time. The same

is true for the measure  $V_2$ , which solely differs from the measure  $V_1$  by taking the square of the summed and averaged distances. The measures  $V_1$ ,  $V_2$  and  $V$  each determine the explained variation by comparing the distances of a singleton survival curve to the survival curve derived from a Cox model with given covariates. The Cox model is based on the assumption of proportional hazards, which therefore should be checked prior to application. This is not required for the novel measure  $V_1^*$ . However, it might be possible to compute the measures  $V_1$ ,  $V_2$  and  $V$  also with other survival models than the Cox model to avoid the assumption of proportional hazards.

Furthermore, the persons who are censored prior to the first event time need to be removed from the dataset prior to the computation of the measures  $V_1$  and  $V_2$ . In these measures the weighting term for a censored person is given by the number of events prior to the censoring of a person, which is 0 for these persons. This results in an impossible division by 0 for the persons, who are censored prior to the first event time. In contrast to this in the novel measure  $V_1^*$  the weighting term is depicted in a different way, with the consequence that the persons, who are censored prior to the first event time, can remain in the dataset.

As mentioned in the introduction, several measures of explained variation for survival data have been proposed in recent years. The measures applied in this work either follow a “distance-based” or a “likelihood-based” approach. In a “distance-based” approach a scoring rule measures the distance between predicted and observed survival outcome. A further popular measure in this field is the Brier score, which is calculated at fixed time points and afterwards combined to give a time-dependent curve [29], [30], [31]. In a “likelihood-based” approach a measure relates the log likelihood of a model with covariate information to the log likelihood obtained from a “null model” ignoring covariate information [29]. Furthermore many measures have been proposed, which follow a “discrimination-based” approach. These measures are popular tools to characterise the predictive performance of a survival model and use this model to distinguish between persons having an event and persons having no event at any specific time point. For example, the time-dependent area under the ROC curve can be used as a measure of the discriminative ability of the survival model at each time point [29], [30].

Further research could be performed for the novel measure  $V_1^*$ , which is proposed in this work. One step would be the construction of a confidence interval for the measure itself as well as the construction of a confidence interval for the mean difference of the explained variation of two covariates. Furthermore it would be important to determine the maximum value for the measure  $V_1^*$ .

The measure  $V_2$  as well as the measure  $R_{LR}^2$  are bounded above by a constant less than 1 [21], [32]. For the measure  $V_1^*$  the maximum value in a dataset is unknown. Theoretically the measure has the maximum value of 1, if the number of categories is equal to the number of



event times in a dataset. This is for example the case if every person is in his or her own category. Furthermore this is the case if several persons are in the same category and these persons experience the event at the exact same time or if some of these persons experience the event at the exact same time and the others are censored. Contrarily, if the number of categories is smaller than the number of event times, at least one category exists in which the persons experience the event at different event times. In this latter case the measure  $V_1^*$  cannot gain the maximum value of 1. Therefore it might be sensible to depict the maximum value of the measure  $V_1^*$  in a given dataset. Afterwards the value of the measure could be standardised to this maximum value. For a limited number of persons the exact maximum value could be identified by a simulation of all possible permutations of the number of events, the number of censorings as well as the number of categories. For a large number of persons an approximation thereof would be feasible. The functions to compute the novel measure  $V_1^*$  and to apply the test of significance can be obtained on request. The circumstance that some measures of explained variation cannot gain the maximum value of 1 might be responsible for the huge range of values in the measures of explained variation in the dataset of persons with histopathological papillary thyroid cancer. Based on this dataset we seek to answer the question whether the fifth, the sixth or the seventh version of the UICC classification system is the “better” classification system for the time to the occurrence of a distant metastasis. A traditional medical approach to answer the question concerning the “best” version of the UICC classification system would be a prospective randomised clinical trial. A person would be randomised to the fifth, the sixth or the seventh version of the UICC classification system and the tumour would be categorised accordingly. At the end of the study the time to the occurrence of the outcome variable could be compared for the versions of the UICC classification system. However, for persons with thyroid cancer, the performance of a prospective randomised clinical trial is not feasible in this context due to different reasons. First of all a differentiated thyroid tumour is a rare disease with an usually slow natural progression [18]. Furthermore different versions of the UICC tumour classification system are not valid at the same time, rather replaces an updated version the previous version.

Therefore, to answer the question which version of the UICC classification system is the “best” version, we compare the versions according to their proportion of explained variation and the “best” version is the version which explains the most variation. In spite of the huge range of values of the measures of explained variation for the persons with histopathological papillary thyroid cancer, all measures have in common that the sixth version explains more variation than the fifth version of the UICC tumour classification system for the outcome variable. The values of the measures of explained variation according to the seventh version are, in comparison to the values of the sixth version, identical or slightly larger.

However, one has to keep in mind, that the fifth and the sixth version each comprise four categories whereas the seventh version consists of five categories. The gain in the values of the measures of explained variation for the seventh version in comparison to the values of the sixth version might be induced by the existence of an additional category in the seventh version. A similar circumstance is known for the coefficient of determination where an increasing number of predictors usually leads to an increase in the coefficient of determination, even when the true values of the new regression coefficients are zero [33]. Therefore, the comparison of the fifth and the sixth version of the UICC tumour classification system might be more appropriate than the comparison of the sixth and the seventh version. It can be stated that the sixth version of the UICC tumour classification system is “better” than the fifth version with respect to the amount of explained variation for the outcome variable “time to the occurrence of a distant metastasis” in the dataset of persons with histopathological papillary thyroid cancer.

## 5 Conclusion

The novel measure of explained variation proposed in this work is a measure with a comprehensible derivation as well as a graphical interpretation. It is based only on the Kaplan-Meier estimator and is therefore a completely nonparametric measure. Furthermore, we propose the application of a statistical test of significance additionally to the computation of the measure itself. The novel measure may for categorical covariates be used as a measure of explained variation in further analyses with survival data and the obtained results might additionally be evaluated by inferential means.

## Data

Data for this article are available from the Dryad Digital Repository: <http://dx.doi.org/10.5061/dryad.5c6bq> [34].

## Notes

### Competing interests

The authors declare that they have no conflicts of interest in the research.

## Appendix

Details for the computation of the measures of explained variation in subsection 3.4 are given below.

The functions to compute the novel measure  $V_1^*$  and to apply the corresponding test of significance are developed for the software package R 2.15.1 [23] and can be obtained on request. The computations of the measures

$R_{LR}^2$  [20] and  $\rho_{W,A}^2$  [22] are directly carried out in SAS® 9.3 (TS1M0) with SAS/STAT 9.3 [24] according to the definition of the measures. For the remaining measures the following programs, functions and options are applied: Lachin [32] provides on the website accompanying the first edition of his book (<http://www.bsc.gwu.edu/bsc/webpage.php?no=18> (last accessed 03/09/2015)) a program for the computation of the measure  $V_2$  by Schemper [2]. This program is subject to minor changes prior to application: the term  $k_i$  is altered to the number of events in analogy to the definition by Schemper, whereas in the program by Lachin the term  $k_i$  is defined as the number of event times. Furthermore the squaring in the measure  $V_2$  is positioned at the correct spot, the covariates are considered as categorical variables in the respective Cox model and the persons who are censored prior to the first event time are removed from the dataset. To compute the measure  $V_1$  by Schemper [2] this program is extended. For the computation of the measures  $V_1$  and  $V_2$  the software package SAS® 9.3 (TS1M0) with SAS/STAT 9.3 [24] is used. The measure  $V$  by Schemper and Henderson [19] is determined with the option “EV” in the procedure “PHREG” in SAS® 9.4 (TS1M1) with SAS/STAT 13.1 [25], the measure  $\bar{R}_{LR}^2$  [21] with the function “cph”, package “rms” in R 2.15.1 [23] and the measure  $\rho_{W}^2$  [22] with the macro “KENTOQNR” by Heinzl [35] in SAS 9.3® (TS1M0) with SAS/STAT 9.3 [24]. For all measures, which are based on the computation of the likelihood function, the approximation by Breslow [36] is used.

## References

- Seber GA, Lee AJ. Linear regression analysis. 2nd ed. Hoboken: John Wiley; 2003. DOI: 10.1002/9780471722199
- Schemper M. The explained variation in proportional hazards regression. *Biometrika*. 1990;77(1):216-8. Erratum in: *Biometrika*. 1994;81(3):631. DOI: 10.1093/biomet/77.1.216
- Barbera L, Groome PA, Mackillop WJ, Schulze K, O'Sullivan B, Irish JC, Warde PR, Schneider KM, Mackenzie RG, Hodson DI, Hammond JA, Gulavita SP, Eapen LJ, Dixon PF, Bissett RJ. The role of computed tomography in the T classification of laryngeal carcinoma. *Cancer*. 2001 Jan 15;91(2):394-407. DOI: 10.1002/1097-0142(20010115)91:2<394::AID-CNCR1014>3.0.CO;2-J
- Mehaffey MG, Folberg R, Meyer M, Bentler SE, Hwang T, Woolson R, Moore KC. Relative importance of quantifying area and vascular patterns in uveal melanomas. *Am J Ophthalmol*. 1997 Jun;123(6):798-809. DOI: 10.1016/S0002-9394(14)71129-8
- Key J, Best N, Joffe M, Jensen TK, Keiding N. Methodological issues in analyzing time trends in biologic fertility: protection bias. *Am J Epidemiol*. 2009 Feb;169(3):285-93. DOI: 10.1093/aje/kwn302
- Rivalan P, Barbraud C, Inchausti P, Weimerskirch H. Combined impacts of longline fisheries and climate on the persistence of the Amsterdam Albatross *Diomedea amsterdamensis*. *IBIS*. 2010;152(1):6-18. DOI: 10.1111/j.1474-919X.2009.00977.x
- Gneiting T, Raftery AE. Strictly proper scoring rules, prediction, and estimation. *J Am Stat Assoc*. 2007;102(477):359-78. DOI: 10.1198/016214506000001437
- Lawless JF, Yuan Y. Estimation of prediction error for survival models. *Stat Med*. 2010 Jan 30;29(2):262-74. DOI: 10.1002/sim.3758
- Schemper M, Stare J. Explained variation in survival analysis. *Stat Med*. 1996 Oct 15;15(19):1999-2012. DOI: 10.1002/(SICI)1097-0258(19961015)15:19<1999::AID-SIM353>3.0.CO;2-D
- Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc*. 1958;53(282):457-81. DOI: 10.1080/01621459.1958.10501452
- Schwertmann NC, Gilks AJ, Cameron J. A simple noncalculus proof that the median minimizes the sum of the absolute deviations. *Am Stat*. 1990;44(1):38-9. DOI: 10.1080/00031305.1990.10475690
- Collett D. Modelling survival data in medical research. 2nd ed. Boca Raton: Chapman & Hall; 2003.
- Tarone RE, Ware J. On distribution-free tests for equality of survival distributions. *Biometrika*. 1977;64(1):156-60. DOI: 10.1093/biomet/64.1.156
- Conover WJ. Practical nonparametric statistics. 3rd ed. New York: John Wiley; 1999.
- Sobin LH, Wittekind C; UICC International Union Against Cancer. TNM classification of malignant tumours. 5th ed. New York: Wiley-Liss; 1997.
- Sobin LH, Wittekind C; UICC International Union Against Cancer. TNM classification of malignant tumours. 6th ed. New York: Wiley-Liss; 2002.
- Sobin LH, Gospodarowicz MK, Wittekind Ch; UICC International Union Against Cancer. TNM classification of malignant tumours. 7th ed. Chichester, Hoboken: Wiley-Blackwell; 2010.
- Meixner M, Hellmich M, Dietlein M, Kobe C, Schicha H, Schmidt M. Disease-free survival in papillary and follicular thyroid carcinoma. Comparison between UICC 5th and 7th classifications of T stage, and the prognostic value of primary tumour size. *Nuklearmedizin*. 2013;52(3):71-80. DOI: 10.3413/Nukmed-0530-12-09
- Schemper M, Henderson R. Predictive accuracy and explained variation in Cox regression. *Biometrics*. 2000 Mar;56(1):249-55. DOI: 10.1111/j.0006-341X.2000.00249.x
- Magee L. R<sup>2</sup> measures based on Wald and likelihood ratio joint significance tests. *Am Stat*. 1990;44(3):250-3. DOI: 10.2307/2685352
- Nagelkerke NJ. A note on a general definition of the coefficient of determination. *Biometrika*. 1991;78(3):691-2. DOI: 10.1093/biomet/78.3.691
- Kent JT, O'Quigley J. Measures of dependence for censored survival data. *Biometrika*. 1988;75(3):525-34. DOI: 10.1093/biomet/75.3.525
- R Development Core Team. R: a language and environment for statistical computing. Wien: R Foundation for Statistical Computing; 2012.
- SAS Institute Inc. SAS/STAT® 9.3 User's guide. Cary: SAS Institute Inc.; 2011.
- SAS Institute Inc. SAS/STAT® 13.1 User's guide. Cary: SAS Institute Inc.; 2013.
- Kalbfleisch JD, Prentice RL. The statistical analysis of failure time data. 2nd ed. Hoboken: Wiley; 2002. DOI: 10.1002/9781118032985
- Machin D, Campbell MJ. Design of studies for medical research. Hoboken: Wiley; 2005. DOI: 10.1002/0470012994

28. National Research Council. The prevention and treatment of missing data in clinical trials. Panel on Handling Missing Data in Clinical Trials. Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: National Academies Press; 2010.
29. Schmid M, Potapov S. A comparison of estimators to evaluate the discriminatory power of time-to-event models. *Stat Med*. 2012 Oct;31(23):2588-609. DOI: 10.1002/sim.5464
30. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, Pencina MJ, Kattan MW. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology*. 2010 Jan;21(1):128-38. DOI: 10.1097/EDE.0b013e3181c30fb2
31. Graf E, Schmoor C, Sauerbrei W, Schumacher M. Assessment and comparison of prognostic classification schemes for survival data. *Stat Med*. 1999;18(17-18):2529-45. DOI: 10.1002/(SICI)1097-0258(19990915/30)18:17/18<2529::AID-SIM274>3.0.CO;2-5
32. Lachin JM. Biostatistical methods: the assessment of relative risks. 2nd ed. Hoboken: Wiley; 2011. DOI: 10.1002/9780470907412.indauth
33. Healy MJR. The use of  $R^2$  as a measure of goodness of fit. *J R Stat Soc Ser A*. 1984;147(4):608-9. DOI: 10.2307/2981848
34. Weiß V, Schmidt M, Hellmich M. Data from: A novel nonparametric measure of explained variation for survival data with an easy graphical interpretation. Dryad Digital Repository. 2015. DOI: 10.5061/dryad.5c6bq
35. Heinzl H. Using SAS to calculate the Kent and O'Quigley measure of dependence for Cox proportional hazards regression model. *Comput Methods Programs Biomed*. 2000 Aug;63(1):71-6. DOI: 10.1016/S0169-2607(00)00073-0
36. Breslow N. Covariance analysis of censored survival data. *Biometrics*. 1974 Mar;30(1):89-99. DOI: 10.2307/2529620

**Corresponding author:**

Prof. Dr. Martin Hellmich  
Institute of Medical Statistics, Informatics and  
Epidemiology, University of Cologne, Kerpener Straße 62,  
50924 Köln, Germany, Phone: +49 221 478 6509/6501,  
Fax: +49 221 478 6520  
martin.hellmich@uni-koeln.de

**Please cite as**

Weiß V, Schmidt M, Hellmich M. A novel nonparametric measure of explained variation for survival data with an easy graphical interpretation. *GMS Ger Med Sci*. 2015;13:Doc18. DOI: 10.3205/000222, URN: urn:nbn:de:0183-0002220

**This article is freely available from**

<http://www.egms.de/en/journals/gms/2015-13/000222.shtml>

**Received:** 2015-06-08

**Revised:** 2015-09-17

**Published:** 2015-10-29

**Copyright**

©2015 Weiß et al. This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 License. See license information at <http://creativecommons.org/licenses/by/4.0/>.