WORKING PAPER SERIES
Division of Biostatistics, Karolinska Institutet, Stockholm, Sweden
April 2020

Defining and modeling the probability of occurrence of events

Matteo Bottai, Andrea Discacciati, and Giola Santoni
Karolinska Institutet, Stockholm, Sweden

Abstract

This paper introduces the event-probability function, a measure of occurrence of an event of interest over time, defined as the instantaneous probability of an event at a given time point conditional on having survived until that point. Unlike the hazard function, the event-probability function defines the instantaneous probability of the event. We explore its properties and interpretation and highlight its connection with other distributions functions. We propose convenient methods for modeling the possible effect of covariates, including flexible proportional-odds models and flexible power-probability models, that allow for censored and truncated observations. We contrast the proposed methods with other popular methods, and discuss the theoretical and computational aspects of parameter estimation. Finally, we evaluate the mortality risk in patients with metastatic renal carcinoma from a randomized clinical trial.

Keywords: Cox regression, hazards, rates, survival analysis

1 Introduction

Measuring the occurrence of events, such as death, over time is a central objective in many areas of science. The survival function and the hazard function uniquely define the time-to-event variable of interest and have long been used in applications. In particular, the interest in the hazard function has much grown since the introduction of methods like Cox regression. While the interpretation of the survival function is intuitive to researchers and lay people alike, that of the hazard function is not. In the scientific literature, hazards are often mistaken for probabilities, and in interpreting study findings the word "risk" is usually preferred to word "hazard". For example, a hazard ratio of 1.2 may be read as an increase of 20% in risk. This practice dates long back and has led many a study to misplaced conclusions. Sutradhar and Austin (2018) recently discussed its relevance and magnitude and pointed out that contrary to published guidelines (Guidelines, 2019) and recommendations (Oakes and Peterson, 2008; Sedgwick, 2012), the mistaken interpretation of hazards is ubiquitous.

This paper introduces the event-probability function, which properly defines the instantaneous probability of an event. Unlike the hazard function, the event-probability function is bounded between zero and one and can be interpreted as risk. In the following sections we explore its properties and interpretation, highlight its connection with other distributions functions, propose convenient strategies for modeling the possible effect of covariates and contrast them with other popular existing methods, and discuss the theoretical and computational aspects of parameter estimation. The proposed event-probability models are applied in the evaluation of the mortality risk in patients from a randomized clinical trial with metastatic renal carcinoma.

2 The event-probability function

We consider a time-to-event variable T with support on the positive real half-line. Let $F(t) = \Pr(T \leq t)$, $S(t) = 1 - F(t)$, $f(t) = dF(t)/dt$, $h(t) = f(t)/S(t)$, and $H(t) = \int h(t)dt = -\log[S(t)]$, indicate the cumulative distribution function, survival function, the probability density function, the hazard function, and the cumulative hazard function, respectively. The functions $F(t)$, $S(t)$, and $f(t)$ are defined over the entire real line, $\mathbb{R}$, while $h(t)$ and $H(t)$ are defined over the set $\{t \in \mathbb{R} : S(t) > 0\}$. For clarity, we defer discussing censoring and truncation to Section 4.

2.1 Definition of the event-probability function

The probability of occurrence of an event over the time interval $(t, t + \delta)$, with $\delta > 0$, conditional on $T > t$, is

$$\Pr[T < t + \delta \mid T > t] = 1 - \frac{S(t + \delta)}{S(t)} \tag{1}$$

defined over the set $\{t \in \mathbb{R} : S(t) > 0\}$. Bottai (2017) defined the average probability of occurrence of the event over the interval $(t, t + \delta)$, conditional on $T > t$, as

$$G(t, t + \delta) = 1 - \left[\frac{S(t + \delta)}{S(t)}\right]^{1/\delta} \tag{2}$$

An heuristic interpretation of the probability defined in equation (2) and that of the following related definitions is given in Section 2.2. Because $S(0) = 1$, the survival function can be written as

$$S(t) = [1 - G(0, t)]^t \tag{3}$$

Bottai (2017) also defined the instantaneous probability as the limit of the average probability over shrinking time intervals

$$g(t) = \lim_{\delta \to 0} G(t, t+\delta) \tag{4}$$

As δ tends to zero, the average probability $G(t, t+\delta)$ tends to the function $g(t)$, whereas the conditional probability in equation (1) tends to zero. Henceforth, we refer to the function $g(t)$ as the *event-probability function*.

To simplify the mathematical expressions in the remainder of this paper, we define the average survival functions and the *survival-probability function* respectively as

$$\bar{G}(t) = 1 - G(t)$$

$$\bar{g}(t) = 1 - g(t)$$

2.2 Interpretation of the event-probability function

If the probability of surviving two days is $S(2) = 0.36$, then from equation (2) the average probability per day is $G(0, 2) = 1 - 0.36^{1/2} = 0.40$. Applied every day, this average daily probability yields the two-day probability $S(2) = [1 - G(0, 2)]^2 = (1 - 0.40)^2 = 0.36$, as per equation (3). The event-probability function, $g(t)$, can be interpreted as the average probability over infinitely short time intervals, or instantaneous probability. For example, if an event occurs with a constant, instantaneous daily probability of 0.40, then $g(t) = 0.40$, for all $t > 0$, and the probability of surviving two days is $\bar{g}(t)^2 = (1 - 0.40)^2 = 0.36$.

An analogy with speed may facilitate understanding. If one travels at the constant, instantaneous speed of 10 miles an hour, one will cover 20 miles in two hours. The constant speed in this example is analogous to a constant instantaneous probability, which applied every instant yields the nominal occurrence probability over any given interval of time.

Bottai (2017) showed that the hazard function can be seen as the limit of the average number of events divided by the cumulative person-time over a shrinking time interval,

$$h(t) = \lim_{\delta \to 0} \frac{S(t) - S(t+\delta)}{\int_t^{t+\delta} S(u)du} \tag{5}$$

In the literature, the hazard function often is defined as

$$h(t) = \lim_{\delta \to 0} \frac{P(t < T \leq t+\delta | T > t)}{\delta} \tag{6}$$

In general, the arguments of the limits in equations (5) and (6) are unequal,

$$\frac{P(t < T \leq t+\delta | T > t)}{\delta} = \frac{S(t) - S(t+\delta)}{\delta S(t)} \neq \frac{S(t) - S(t+\delta)}{\int_t^{t+\delta} S(u)du}$$

because $\delta S(t) \neq \int_t^{t+\delta} S(u)du$. Nevertheless, the hazard function can be defined using either of the limits in equations (5) and (6). Note that the arithmetic mean conditional probability $P(t < T \leq t+\delta | T > t)/\delta = [1 - P(T > t+\delta | T > t)]/\delta$, whose limit is used in the definition of the hazard function in equation (6), differs from the geometric average probability in equation (2), $1 - [S(t+\delta)/S(t)]^{1/\delta} = 1 - P(T > t+\delta | T > t)^{1/\delta}$, whose limit is used in the definition of the event-probability function in equation (4). The geometric mean is more appropriate than the arithmetic mean in summarizing the probability of events (Bottai, 2017) and it has long been used in science (among others, Gompertz, 1825).

Bottai (2017) also showed that

$$g(t) = 1 - \exp[-h(t)] \tag{7}$$

The above equality implies the following corollary, whose proof is given in Appendix A.

Corollary 1. *Over the support of the time-to-event variable T, $\mathcal{T} = \{t \in \mathbb{R} : f(t) > 0\}$,*

the event-probability function is strictly smaller than the hazard function, $g(t) < h(t)$.

The above corollary implies that the event-probability function is never equal to the hazard function. The former can be interpreted as an instantaneous probability, while the latter should never be interpreted as such. To further understanding of the difference between these two functions, consider the following two identities

$$g(t) = 1 - \exp[-h(t)]$$
$$F(t) = 1 - \exp[-H(t)]$$

The above identities show that hazard is to the event-probability what the cumulative hazard is to the cumulative distribution. The event-probability and the cumulative distribution are probabilities, and the hazard and the cumulative hazard are not.

The differing interpretation of the event-probability and the hazard is consequential. For example, suppose the probability of an event to occur in a month is equal to $g_0(t) = 0.5$ in a population and $g_1(t) = 0.9$ in another population. The event-probability ratio, or simply risk ratio, is

$$\mathrm{RR}(t) = \frac{g_1(t)}{g_0(t)} \approx 1.8$$

and the event-odds ratio, or simply odds ratio, is

$$\mathrm{OR}(t) = \frac{g_1(t)/\bar{g}_1(t)}{g_0(t)/\bar{g}_0(t)} \approx 9.0$$

Because the event-probabilities can be interpreted as risks, their ratio can be interpreted as the relative risk, or risk ratio. From equation (7), the corresponding hazard functions are equal to $h_0(t) = -\log(1 - 0.5) \approx 0.7$ and $h_1(t) = -\log(1 - 0.9) \approx 2.3$, respectively. These measures should not be interpreted as risks, and therefore their ratio, $2.3/0.7 \approx 3.3$, should not be interpreted as a risk ratio.

3 Modeling event-probabilities

The average probability $G(t, t+\delta)$ and the event-probability $g(t)$ define the distribution of the random variable T. The average probability function is related to the other distribution functions through equation (2) as follows

$$F(t) = 1 - \bar{G}(0,t)^t$$

$$S(t) = \bar{G}(0,t)^t$$

$$f(t) = -\bar{G}(0,t)^t \left\{ t\bar{G}'(0,t)/\bar{G}(0,t) + \log[\bar{G}(0,t)] \right\}$$

$$h(t) = -t\bar{G}'(0,t)/\bar{G}(0,t) - \log[\bar{G}(0,t)]$$

$$H(t) = -t \log \bar{G}(0,t)$$

where $\bar{G}'(0,t)$ indicates the first derivative of $\bar{G}(0,t)$ with respect to t, $\bar{G}'(0,t) = \partial \bar{G}(0,t)/\partial t$.

The event-probability function is related to the other distribution functions through equation (7) as follows

$$F(t) = 1 - \exp\left\{ \int_0^t \log[\bar{g}(u)] du \right\} \tag{8}$$

$$S(t) = \exp\left\{ \int_0^t \log[\bar{g}(u)] du \right\} \tag{9}$$

$$f(t) = -\log[\bar{g}(t)] \exp\left\{ \int_0^t \log[\bar{g}(u)] du \right\} \tag{10}$$

$$h(t) = -\log[\bar{g}(t)] \tag{11}$$

$$H(t) = -\int_0^t \log[\bar{g}(u)] du \tag{12}$$

The following are examples of known distributions specified by their event-probability functions: $\bar{g}(t|\theta) = \exp(-1/\theta)$ defines an exponential distribution with scale parameter $\theta > 0$; $\bar{g}(t|\theta, \eta) = \exp[-(t/\theta)^{\eta-1}\eta/\theta]$ defines a Weibull distribution with scale parameter $\theta > 0$ and shape parameter $\eta > 0$; $\bar{g}(t|\theta, \eta) = \exp[-\theta\eta \exp(\theta t)]$ defines a Gompertz distribution with scale parameter $\theta > 0$ and shape parameter $\eta > 0$; finally, if

$\bar{g}[\log(t|\theta,\mu)] = \exp\{-\exp[(\log(t)-\mu)/\theta]/\theta\}$, then $\log(T)$ follows a Gumbel distribution with location parameter $\mu \in \mathbb{R}$ and scale parameter $\theta > 0$.

Throughout this paper, any parameter can be assumed to depend on a q-dimensional vector of covariates $X \in \mathbb{R}^q$. For example, a positive scale parameter can be defined as $\theta = \exp(X'\beta)$ for a parameter vector $\beta \in \mathbb{R}^q$.

The following lemma shows the relationship between the event-probability function and the product-limit definition of the survival function. The proof is given in the Appendix.

Lemma 1. *Let $u_1 < u_2 < \ldots < u_{k+1}$ be an ordered set of distinct values from $u_1 = 0$ to $u_{k+1} = t$, forming k intervals of length $\delta_k = u_{i+1} - u_i$, with $i = 1, \ldots, k$. From equation (9), assuming $\log[\bar{g}(u)]$ is Riemann-integrable,*

$$S(t) = \lim_{k\to\infty} \prod_{i=1}^{k} \bar{g}(u_i)^{\delta_k}$$

As previously shown by Bottai (2017), the above result is related to the Kaplan-Meier estimator of the survival function, where the event-probability are the steps of the Kaplan-Meier curve. The value $\bar{g}(t)^{\delta_k}$ is an approximation of $G(t, t+\delta)$ over the finite interval $(t, t+\delta)$.

The following corollary of Lemma 1 gives another expression for the cumulative hazard function that may be used in the likelihood discussed in Section 4. The proof follows directly from that of Lemma 1 and is omitted.

Corollary 2. *Under the conditions stated in Lemma 1,*

$$H(t) = -\log[S(t)] = \lim_{k\to\infty} \sum_{i=1}^{k} \delta_k \log[\bar{g}(u_i)]$$

The following lemma gives few more expressions for the cumulative hazard function that may inform modeling strategies. The proof is given in the Appendix.

Lemma 2. The cumulative hazard can be written as

$$\begin{aligned} H(t) &= -t\log[\bar{g}(t)] + \int_0^t \frac{u\bar{g}'(u)}{\bar{g}(u)}du \\ H(t) &= \int_{\bar{g}(0)}^{\bar{g}(t)} \bar{g}^{-1}(u)/u \; du - \left[\log(u)\bar{g}^{-1}(u)\right]\Big|_{\bar{g}(0)}^{\bar{g}(t)} \\ H(t) &= \int_{\bar{g}(0)}^{\bar{g}(t)} -\frac{\log(u)}{\bar{g}'[\bar{g}^{-1}(u)]}du \end{aligned}$$

The following Sections 3.1 and 3.2 describe two convenient classes of models, the proportional-event-odds model and the power-probability model, respectively.

3.1 Proportional-odds model

Because the event-probability function is a probability, and as such bounded between zero and one, it is natural to start by considering proportional-odds models. These can be defined as

$$\frac{g(t|\theta)}{1-g(t|\theta)} = \frac{g_0(t)}{1-g_0(t)}\theta \tag{13}$$

The baseline event-probability function is indicated by $g_0(t)$ and the proportionality parameter by $\theta > 0$. This model corresponds to the survival-probability function

$$\bar{g}(t|\theta) = \left[\frac{g_0(t)}{1-g_0(t)}\theta + 1\right]^{-1}$$

and to the hazard function

$$h(t|\theta) = \log\left[\frac{g_0(t)}{1-g_0(t)}\theta + 1\right]$$

Note that the proportional-odds model in equation (13) is different from other models that in the literature are also referred to as proportional-odds models (Bennett, 1983; Kirmani and Gupta, 2001), which replace the event-probability function $g(t|\theta)$ with the survival function $S(t|\theta)$.

We first define the simplest proportional-odds model, which assumes that the baseline odds are a linear function of time,

$$\frac{g(t|\theta)}{1-g(t|\theta)} = \theta t$$

The above model implies the following distribution functions, which do not belong to any known family,

$$\bar{g}(t|\theta) = (\theta t + 1)^{-1}$$

$$h(t|\theta) = \log(\theta t + 1)$$

$$H(t|\theta) = (1/\theta + t)\log(\theta t + 1) - t$$

$$S(t|\theta) = \exp(t)(\theta t + 1)^{-1/\theta - t}$$

$$f(t|\theta) = \exp(t)(\theta t + 1)^{-1/\theta - t}\log(\theta t + 1)$$

We now consider a model with nonlinear baseline odds functions

$$\frac{g(t|\theta)}{1-g(t|\theta)} = \theta t^{\eta}$$

which implies the following distribution functions, which do not belong to any known family,

$$\bar{g}(t) = (\theta t^{\eta} + 1)^{-1}$$

$$h(t) = \log(\theta t^{\eta} + 1)$$

$$H(t) = t[{}_2F_1(-\theta t^{\eta})\eta + \log(\theta t^{\eta} + 1) - \eta]$$

$$S(t) = \exp\{-t[{}_2F_1(-\theta t^{\eta})\eta + \log(\theta t^{\eta} + 1) - \eta]\}$$

$$f(t) = \log(\theta t^{\eta} + 1)\exp\{-t[{}_2F_1(-\theta t^{\eta})\eta + \log(\theta t^{\eta} + 1) - \eta]\}$$

where ${}_2F_1$ indicates the generalized hypergeometric function

$${}_2F_1(-\theta t^{\eta}) = \sum_{i=0}^{\infty} \frac{(1/\eta)_i}{(1 + 1/\eta)_i} \frac{(-\theta t^{\eta})^i}{i!}$$

with $(a)_i = 1$ if $i = 0$ and $(a)_i = a(a+1)(a+2)\cdots(a+i-1)$ if $i > 0$.

Finally, we consider a convenient flexible parametric model. The proportional-odds model defined in equation (13) can be written through a logarithmic transform as

$$\log\left[\frac{g(t|\theta)}{1 - g(t|\theta)}\right] = \log\left[\frac{g_0(t)}{1 - g_0(t)}\right] + \log(\theta)$$

The log-odds transform of the baseline function $g_0(t)$ can be modeled through flexible parametric functions

$$\log\left[\frac{g(t|\theta, \eta)}{1 - g(t|\theta, \eta)}\right] = s(t)'\eta + \log(\theta) \tag{14}$$

where η represents an r-dimensional parameter vector and

$$s(t) = [s_1(t), \ldots, s_r(t)]' \tag{15}$$

is a basis of r functions of t, such as Legendre polynomials or regression cubic splines. Because of the log-odds transform on the left-hand side of equation (14) can take on

values on the entire real line, the expression on the right-hand side is unconstrained. This confers model (14) desirable flexibility, as illustrated in the real-data analysis presented in Section 5. The use of flexible parametric models has been advantageously utilized in many other settings (among others, Royston and Parmar, 2002).

3.2 Power-probability (proportional-hazard) model

This section describes power-probability models, as a possible alternative to the proportional-odds models presented in Section 3.1. A power-probability model is defined as

$$\bar{g}(t|\theta) = \bar{g}_0(t)^{\theta} \tag{16}$$

for a baseline average survival function $\bar{g}_0(t)$ and power parameter $\theta > 0$. If $\theta = 1$, then $g(t) = g_0(t)$. If it exists, the k-th derivative, $k \in \mathbb{N}$, of $\bar{g}(t|\theta)$ with respect to θ is

$$\frac{d^k}{d\theta^k}\bar{g}(t|\theta) = \bar{g}(t|\theta)\log[g_0(t)]^k$$

Because $\log[g_0(t)]$ is negative, the above expression indicates that the event-probability function is decreasing and convex with respect to the parameter θ with the following limit behavior

$$\lim_{\theta\to 0} g(t) = 1 - \lim_{\theta\to 0}\bar{g}(t) = 0$$
$$\lim_{\theta\to \infty} g(t) = 1 - \lim_{\theta\to \infty}\bar{g}(t) = 1$$

Larger values of θ correspond to larger values of $g(t|\theta)$ and therefore larger probabilities of the event. The power-probability model defined in equation (16) corresponds to a proportional-hazard model

$$h(t|\theta) = -\log[\bar{g}(t|\theta)] = -\log[\bar{g}_0(t)]\theta = h_0(t)\theta$$

with baseline hazard function $h_0(t) = -\log[\bar{g}_0(t)]$ and proportionality parameter $\theta > 0$. Traditional parametric and semi-parametric models can be used to estimate the parameter θ. A hazard ratio θ corresponds to the following event-probability

$$g(t|\theta) = 1 - \bar{g}_0(t|\theta)^\theta \tag{17}$$

to the following risk ratio

$$\mathrm{RR}(t|\theta) = \frac{g(t|\theta)}{g_0(t|\theta)} = \frac{1 - \bar{g}_0(t|\theta)^\theta}{1 - \bar{g}_0(t|\theta)} \tag{18}$$

and to the following odds ratio

$$\mathrm{OR}(t|\theta) = \frac{g(t|\theta)/[1 - g(t|\theta)]}{g_0(t|\theta)/[1 - g_0(t|\theta)]} = \frac{[1 - \bar{g}_0(t|\theta)^\theta]/\bar{g}_0(t|\theta)^\theta}{[1 - \bar{g}_0(t|\theta)]/\bar{g}_0(t|\theta)} \tag{19}$$

Table 6 shows $g(t|\theta)$, $\mathrm{RR}(t|\theta)$, and $\mathrm{OR}(t|\theta)$, for different baseline average probabilities $g_0(t|\theta)$, corresponding baseline hazards $h_0(t|\theta)$, and probability powers (hazard ratios) θ.

The power-probability model in equation (16) can be written through a complementary-log-log model for the average survival function

$$\log\{-\log[\bar{g}(t|\theta)]\} = \log\{-\log[\bar{g}_0(t)]\} + \log(\theta)$$

which corresponds to a logarithmic transform of the hazard function $\log[h(t|\theta)] = \log[h_0(t)] + \log(\theta)$. As for the proportional-odds model described in Section 3.1, the baseline complementary-log-log function can be modeled through flexible parametric functions

$$\log\{-\log[\bar{g}(t|\theta, \eta)]\} = s(t)'\eta + \log(\theta) \tag{20}$$

where η represents an r-dimensional parameter vector and $s(t)$ is defined in equation (15). Similarly to the log-odds transform in equation (14), the complementary-log-log transform allows the expression on the right-hand side of equation (20) to take on values on the entire real line $\mathbb{R}$, which can make this model practical in many applications. We apply the power-probability model in equation (20) in the real-data analysis reported in Section 5.

4 Estimation and computation

In addition to the notation defined in Section 2, let C be a right-censoring random variable and Z a left-truncating random variable, both with support on the positive real half-line. Let $X \in \mathbb{R}^q$ be a q-dimensional vector of covariates. We can observe the random vector (Y, Z, D, X) only if $Y > Z$, where $Y = \min(T, C)$ and $D = I(T \leq C)$.

We consider a random sample of n covariate vectors $x_1, \ldots, x_n$, possibly right-censored and left-truncated observations $y_1, \ldots, y_n$, left-truncating observations $z_1, \ldots, z_n$, and event indicators $d_1, \ldots, d_n$. The log-likelihood function can be defined through the conditional hazard functions given the covariates and a parameter vector θ as

$$l_n(\theta) = \sum_{i=1}^{n} d_i \log[h(y_i|x_i, \theta)] - H(y_i|x_i, \theta) + H(z_i|x_i, \theta) \tag{21}$$

The above log-likelihood function can also be written under any specification of $g(y_i|x_i, \theta)$ through equations (11) and (12) as

$$l_n(\theta) = \sum_{i=1}^{n} d_i \log\{-\log[\bar{g}(y_i|x_i, \theta)]\} + \int_0^{y_i} \log[\bar{g}(u|x_i, \theta)]du - \int_0^{z_i} \log[\bar{g}(u|x_i, \theta)]du \tag{22}$$

For example, if we assume the proportional-odds model defined in equation (13) with $\theta = \exp(X'\beta)$,

$$\frac{g(t|x_i, \beta)}{1 - g(t|x_i, \beta)} = \exp(x_i'\beta)t \tag{23}$$

then the log-likelihood function in equation (21) has closed-form expression with

$$h(t|x_i, \beta) = \log[\exp(x_i'\beta)t + 1] \tag{24}$$

$$H(t|x_i, \beta) = [\exp(-x_i'\beta) + t] \log[\exp(x_i'\beta)t + 1] - t \tag{25}$$

When the integral in equation (12) does not have closed-form expression, it can be numerically approximated.

If it is of interest to define a parametric model $\bar{G}(0,t|\theta)$ for the average-probability function defined in equation (2), instead of the probability-event function, the log-likelihood function is

$$l_n(\theta) = \sum_{i=1}^{n} d_i \log\left\{ \frac{\bar{G}'(0,y_i|\theta)}{G(0,y_i|\theta)} y_i + \log[\bar{G}(0,y_i|\theta)] \right\} - y_i \log[\bar{G}(0,y_i|\theta)] + z_i \log[\bar{G}(0,z_i|\theta)] \tag{26}$$

The maximum likelihood estimator for θ is defined as the maximizer of $l_n(\theta)$, defined in equation (21) or equation (26), over a parameter space Θ

$$\hat{\theta}_n = \arg\max_{\theta\in\Theta} l_n(\theta) \tag{27}$$

Inference on θ can be made within the theoretical framework of maximum likelihood estimation, which makes $\hat{\theta}_n$ an efficient and practical estimator. The following two theorems state its consistency and asymptotic normality, respectively. Their proofs follows standard likelihood theory, and we omit them for brevity.

Theorem 1 *(Consistency). If the variables $t_1, \ldots, t_n$ are independent and identically distributed as $f(t|\theta)$ in equation (10), the parameter space Θ is compact, the true value θ_0 is identified $\theta \neq \theta_0 \Leftrightarrow f(t|\theta) \neq f(t|\theta_0)$, $\theta_0 = \arg\max_{\theta\in\Theta} E_{\theta_0} l_n(\theta)$, the likelihood function is continuous in θ, $E_{\theta_0} l_n(\theta)$ exists, the likelihood function is such that $l_n(\theta)/n$ converges in probability to $E_{\theta_0} l_n(\theta)$ uniformly in $\theta \in \Theta$, i.e. $\lim_{n\to\infty} P(\sup_{\theta\in\Theta} |l_n(\theta)/n - E_{\theta_0} l_n(\theta)| < \epsilon) = 1$ for any $\epsilon > 0$, then the sequence $\hat{\theta}_n$ converges in probability to θ_0.*

Theorem 2 *(Asymptotic normality). If in addition to the assumptions stated in Theorem 1, θ is in the interior of the parameter space Θ, the likelihood function is twice differentiable in a neighborhood of θ_0, integration and differentiation are interchangeable,*

the information matrix $I(\theta_0) = E_{\theta_0}(-\partial^2 l_n(\theta)/\partial\theta\partial\theta'|_{\theta_0})$ *exists and is non-singular, then the sequence* $\sqrt{n}(\hat{\theta}_n - \theta_0)$ *converges in distribution to a normal variable with mean zero and variance* $I^{-1}(\theta_0)$.

Given a set of data, the value of $\hat{\theta}$ can be computed with maximization functions implemented in popular software programs, such as R, Matlab, Stata, and SAS. These programs also allow numerical evaluation of the integral in equation (12), when this does not have closed-form. The Stata code used for the analyses reported in Section 5 is available in the online supplementary material.

5 Survival in metastatic renal carcinoma patients

In this section, we use proportional-odds models and power-probability models in the evaluation of survival in patients with metastatic renal carcinoma. The data arose from a multi-center randomized controlled trial with 350 patients, assigned to either subcutaneous interferon-α (IFN) or oral medroxyprogesterone (MPA) between 1992 and 1997. The primary endpoint was overall mortality. Three patients had no follow-up data. By June 2001, 322 patients died (93%). The median follow-up time was approximately seven months. A detailed description of patients characteristics, treatment, and follow-up can be found in Medical Research Council Renal Cancer Collaborators (1999) and Royston et al. (2004).

All the models were estimated by maximizing the likelihood function with the user-written `stpreg` command in Stata, a simple adaptation of the more general `stgenreg` command (Crowther and Lambert, 2013). The `stpreg` command improves on the earlier approach proposed by Discacciati and Bottai (2017), in that it does not require splitting of

the data into risk sets or approximating the event-function on a discrete set of time points. The code to reproduce the results is available in the online supplementary material.

5.1 Proportional-odds model

In this subsection we estimate four, increasingly complex, proportional-odds models. The estimates are shown in Table 2. We started by modeling the odds of death for the two treatment groups with a proportional-odds model that assumed constant event-odds throughout the follow-up time,

$$\log\left[\frac{g(t|\eta,\theta)}{1-g(t|\eta,\theta)}\right] = \eta_0 + \theta_1 \text{trt} \tag{28}$$

where the name "trt" indicates the binary treatment indicator that takes on value 0 for the patients randomized to MPA and 1 for those randomized to IFN. The estimated mortality odds was equal to $\exp(\hat{\eta}_0) = 1.79$ in MPA patients and to $\exp(\hat{\eta}_0 + \hat{\theta}_1) = 1.07$ in IFN patients. The mortality odds ratio (OR) was $\exp(\hat{\theta}_1) = 0.60$ with a 95% confidence interval (95% CI) equal to $(0.43, 0.83)$. The model assumed that the OR was constant over the entire duration of the follow-up. Under this model, the odds of death in the IFN group was estimated to be 40% lower than that in the MPA group.

We relaxed the assumption of constant mortality odds over the follow-up period and let the log-mortality odds be a linear function of $\log(t)$ with equal slope for the two treatment groups,

$$\log\left[\frac{g(t|\eta,\theta)}{1-g(t|\eta,\theta)}\right] = \eta_0 + \eta_1 \log(t) + \theta_1 \text{trt} \tag{29}$$

The estimate for the OR comparing the two treatment groups was equal to $\exp(\hat{\theta}_1) = 0.63$ (95% CI: 0.45, 0.87). Under this model, we estimated a $(1 - \exp(\hat{\eta}_1)) \times 100 = 18\%$ decrease in the mortality odds for every one log-year increase in follow-up time in both

treatment groups.

We then relaxed also the linearity assumption by applying a restricted cubic splines transform (RCS) with two degrees of freedom to log-time. The three knots were placed at the minimum, median and maximum of the observed distribution of the uncensored log-time values. We let $s_1(\log(t)) \equiv \log(t)$ and $s_2(\log(t))$ be the first and second RCS transform, respectively. The model was

$$\log\left[\frac{g(t|\eta,\theta)}{1-g(t|\eta,\theta)}\right] = \eta_0 + \eta_1 \log(t) + \eta_2 s_2(\log(t)) + \theta_1 \text{trt} \tag{30}$$

The probability of dying predicted by the model defined in equation (30) is displayed in Figure 1, panel C. After a steep increase in the first 3 months after randomization, the probability in MPA patients decreased from about 0.7 to 0.3 during the follow-up period. The OR was estimated to be $\exp(\hat{\theta}_1) = 0.63$ (95% CI: 0.45, 0.89). The model defined in equation (29) was nested within that defined in equation (30), as the former was equal to the latter when η_2 is equal to zero. Testing the null hypothesis $H_0 : \eta_2 = 0$ helped select the better-fitting model, with small p-values supporting the latter model. In our data, the p-value from the Wald's test was less than 0.001, suggesting a better fit of model defined in equation (30).

Finally, we relaxed the assumption of proportionality in the odds between IFN and MPA treatment groups, by including the interaction (product) terms between the two RCS covariates and the binary treatment indicator.

$$\log\left[\frac{g(t|\eta,\theta)}{1-g(t|\eta,\theta)}\right] = \eta_0 + \eta_1 \log(t) + \eta_2 s_2(\log(t)) + \theta_1 \text{trt} + \theta_2 \log(t)\text{trt} + \theta_3 s_2(\log(t))\text{trt} \tag{31}$$

The model defined in equation (30) was nested within that in equation (31), as the latter was equal to the former when $\theta_2 = \theta_3 = 0$. The p-value from the two-degree-of-freedom

Wald's test for the hypothesis that the coefficients of the two interaction terms were jointly equal to zero was equal to 0.55, providing insufficient evidence of a time-varying OR.

5.2 Power-probability (proportional-hazards) model

We repeated the steps taken in the previous section, this time using power-probability models. The estimates are shown in Table 3. We started by including the binary treatment indicator as the only covariate, thus assuming a constant event-probability throughout the follow-up time. The model was

$$\log\{-\log\left[1 - g(t|\eta, \theta)\right]\} = \eta_0 + \theta_1 \text{trt} \tag{32}$$

The estimated probability of dying was $1 - \exp(-\exp(\hat{\eta}_0)) = 0.64$ on MPA and $1 - \exp(-\exp(\hat{\eta}_0 + \hat{\theta}_1)) = 0.52$ on IFN. The power parameter was estimated to be $\exp(\hat{\theta}_1) = 0.71$ (95% CI: 0.57, 0.88), which meant that the survival-probability at any given time point of the follow-up in MPA patients was equal to that in IFN patients raised by a power of 0.71. Because the power was smaller than one, IFN patients had a smaller mortality risk than MPA patients. The parameter θ_1 can also be interpreted as the hazard ratio for IFN versus MPA patients. The model defined in equation (32) implies an exponential distribution for T with parameter $\lambda = \exp(\eta_0 + \theta_1 \text{trt})$.

Next, we let the complementary log-log transform of death probability vary over time as a linear function of $\log(t)$ with the following power-probability model,

$$\log\{-\log\left[\bar{g}(t|\eta, \theta)\right]\} = \eta_0 + \eta_1 \log(t) + \theta_1 \text{trt} \tag{33}$$

The estimated power parameter was equal to $\exp(\hat{\theta}_1) = 0.73$ (95% CI: 0.58, 0.91), again indicating a better survival probability in IFN patients. By using equation (11), the model

defined in equation (33) can be shown to correspond to Weibull proportional-hazards model with a non-standard parametrization. To see this, let us consider the log-hazard function $\log(\tilde{h}(t|\upsilon, \gamma_0, \gamma_1))$ of a Weibull distribution with shape parameter υ and scale parameter $\lambda = \exp(\gamma_0 + \gamma_1 \text{trt})$,

$$\log(\tilde{h}(t|\upsilon, \lambda_0, \lambda_1)) = \gamma_0 + \gamma_1 \text{trt} + \log(\upsilon) + (\upsilon - 1)\log(t) \tag{34}$$

The right-hand side of equation (33) is equal to the right-hand side of equation (34) with the alternative parameterization $\eta_0 = \gamma_0 + \log(\upsilon)$, $\eta_1 = (\upsilon - 1)$, and $\theta_1 = \gamma_1$.

Next, we relaxed the assumption that the complementary log-log transform of the event-probability was a linear function of $\log(t)$, by including the two RCS covariates $s_1(\log(t)) \equiv \log(t)$ and $s_2(\log(t))$ introduced in the previous section. The power-probability model was

$$\log\{-\log[\bar{g}(t|\eta, \theta)]\} = \eta_0 + \eta_1 \log(t) + \eta_2 s_2(\log(t)) + \theta_1 \text{trt} \tag{35}$$

The Wald's test for $H_0 : \eta_2 = 0$ gave a p-value less than 0.001, providing evidence of non-linearity on the complementary log-log scale. We therefore concluded that the model defined in equation (35) fitted the data better than that in equation (33). The power parameter indicated a better survival in the IFN treatment arm ($\exp(\hat{\theta}_1) = 0.74$ (95% CI: 0.60, 0.92).

Lastly, we allowed the power parameter to vary over time by including interaction terms between the two RCS covariates and the binary treatment indicator. The model was

$$\log\{-\log[\bar{g}(t|\eta, \theta)]\} = \eta_0 + \eta_1 \log(t) + \eta_2 s_2(\log(t)) + \theta_1 \text{trt} + \theta_2 \log(t)\text{trt} + \theta_3 s_2(\log(t))\text{trt}. \tag{36}$$

The p-value from the Wald test for the hypothesis $H_0 : \theta_2 = \theta_3 = 0$ was 0.57, and we concluded that model (36) did not significantly improve the fit of the data over model (35).

We compared the goodness of fit of all the proportional-odds models and power-probability models with the Akaike Information Criterion (AIC). The proportional-odds model defined in equation (30) showed the smallest value (AIC = 1135.8). Figure 1 displays the survival, density, event-probability, and hazard functions implied by this model.

6 Final remarks

This paper shows that modeling and estimation with the event-probability function is as simple as it is with the hazard function. Unlike the hazard function, however, the event-probability function can directly be interpreted as a risk function. Using the latter instead of the former can help redress the acknowledged, ubiquitous misinterpretations that have long afflicted applied research.

SUPPLEMENTARY MATERIAL

"Analysis.do" file: The Stata do-file that produces all the results presented in Section 5, Tables 2 and 3, and Figure 1.

References

Bennett, S. (1983). Analysis of survival data by the proportional odds model. *Statistics in Medicine 2*(2), 273–277.

Bottai, M. (2017). A regression method for modelling geometric rates. *Statistical Methods in Medical Research 26*(6), 2700–2707.

Crowther, M. and P. Lambert (2013). stgenreg: A stata package for general parametric survival analysis. *Journal of Statistical Software 53*(12), 1–17.

Discacciati, A. and M. Bottai (2017). Instantaneous geometric rates via generalized linear models. *Stata Journal 17*(2), 358–371.

Gompertz, B. (1825). On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. *Philosophical Transactions of the Royal Society of London 115*, 513–583.

Guidelines (2019). Information for authors, general statistical guidance. *Annals of Internal Medicine*, http://annals.org/aim/pages/AuthorInformationStatisticsOnly.

Kirmani, S. N. U. A. and R. C. Gupta (2001, Jun). On the proportional odds model in survival analysis. *Annals of the Institute of Statistical Mathematics 53*(2), 203–216.

Medical Research Council Renal Cancer Collaborators (1999). Interferon-alpha and survival in metastatic renal carcinoma: early results of a randomised controlled trial. *Lancet (London, England) 353*(9146), 14–17.

Oakes, D. and D. R. Peterson (2008). Survival methods. *Circulation 117*(22), 2949–2955.

Royston, P. and M. K. B. Parmar (2002). Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Statistics in Medicine 21*(15), 2175–2197.

Royston, P., W. Sauerbrei, and A. Ritchie (2004). Is treatment with interferon-alpha effective in all patients with metastatic renal carcinoma? A new approach to the investigation of interactions. *British Journal of Cancer 90*(4), 794–799.

Sedgwick, P. (2012). Hazards and hazard ratios. *BMJ 345*, e5980.

Sutradhar, R. and P. C. Austin (2018). Relative rates not relative risks: addressing a widespread misinterpretation of hazard ratios. *Annals of Epidemiology 28*(1), 54 – 57.

Table 1: Event-probability $g(t|\theta)$ from equation (17), risk ratio, $\mathrm{RR}(t|\theta)$ from equation (18), and odds ratio, $\mathrm{OR}(t|\theta)$ from equation (19), for different baseline average probabilities $g_0(t|\theta)$, corresponding baseline hazards $h_0(t|\theta)$ from equation (11), and probability powers (hazard ratios) equal to θ.

			Probability Power (Hazard Ratio)						
$g_0(t\|\theta)$	$h_0(t\|\theta)$		0.1	0.5	0.8	1.0	1.3	2.0	3.0
0.001	0.001	$g_1(t\|\theta)$	0.000	0.001	0.001	0.001	0.001	0.002	0.003
		$\mathrm{RR}(t\|\theta)$	0.100	0.500	0.800	1.000	1.300	1.999	2.997
		$\mathrm{OR}(t\|\theta)$	0.100	0.500	0.800	1.000	1.300	2.001	3.003
0.005	0.005	$g_1(t\|\theta)$	0.001	0.003	0.004	0.005	0.006	0.010	0.015
		$\mathrm{RR}(t\|\theta)$	0.100	0.501	0.800	1.000	1.299	1.995	2.985
		$\mathrm{OR}(t\|\theta)$	0.100	0.499	0.800	1.000	1.301	2.005	3.015
0.010	0.010	$g_1(t\|\theta)$	0.001	0.005	0.008	0.010	0.013	0.020	0.030
		$\mathrm{RR}(t\|\theta)$	0.100	0.501	0.801	1.000	1.298	1.990	2.970
		$\mathrm{OR}(t\|\theta)$	0.100	0.499	0.799	1.000	1.302	2.010	3.030
0.050	0.051	$g_1(t\|\theta)$	0.005	0.025	0.040	0.050	0.065	0.098	0.143
		$\mathrm{RR}(t\|\theta)$	0.102	0.506	0.804	1.000	1.290	1.950	2.853
		$\mathrm{OR}(t\|\theta)$	0.098	0.494	0.796	1.000	1.310	2.053	3.161
0.100	0.105	$g_1(t\|\theta)$	0.010	0.051	0.081	0.100	0.128	0.190	0.271
		$\mathrm{RR}(t\|\theta)$	0.105	0.513	0.808	1.000	1.280	1.900	2.710
		$\mathrm{OR}(t\|\theta)$	0.095	0.487	0.791	1.000	1.321	2.111	3.346
0.300	0.357	$g_1(t\|\theta)$	0.035	0.163	0.248	0.300	0.371	0.510	0.657
		$\mathrm{RR}(t\|\theta)$	0.117	0.544	0.827	1.000	1.237	1.700	2.190
		$\mathrm{OR}(t\|\theta)$	0.085	0.456	0.770	1.000	1.376	2.429	4.469
0.500	0.693	$g_1(t\|\theta)$	0.067	0.293	0.426	0.500	0.594	0.750	0.875
		$\mathrm{RR}(t\|\theta)$	0.134	0.586	0.851	1.000	1.188	1.500	1.750
		$\mathrm{OR}(t\|\theta)$	0.072	0.414	0.741	1.000	1.462	3.000	7.000
0.700	1.204	$g_1(t\|\theta)$	0.113	0.452	0.618	0.700	0.791	0.910	0.973
		$\mathrm{RR}(t\|\theta)$	0.162	0.646	0.883	1.000	1.130	1.300	1.390
		$\mathrm{OR}(t\|\theta)$	0.055	0.354	0.694	1.000	1.621	4.333	15.444
0.900	2.303	$g_1(t\|\theta)$	0.206	0.684	0.842	0.900	0.950	0.990	0.999
		$\mathrm{RR}(t\|\theta)$	0.229	0.760	0.935	1.000	1.055	1.100	1.110
		$\mathrm{OR}(t\|\theta)$	0.029	0.240	0.590	1.000	2.106	11.000	>100
0.950	2.996	$g_1(t\|\theta)$	0.259	0.776	0.909	0.950	0.980	0.997	1.000
		$\mathrm{RR}(t\|\theta)$	0.272	0.817	0.957	1.000	1.031	1.050	1.052
		$\mathrm{OR}(t\|\theta)$	0.018	0.183	0.526	1.000	2.533	21.000	>100
0.990	4.605	$g_1(t\|\theta)$	0.369	0.900	0.975	0.990	0.997	1.000	1.000
		$\mathrm{RR}(t\|\theta)$	0.373	0.909	0.985	1.000	1.008	1.010	1.010
		$\mathrm{OR}(t\|\theta)$	0.006	0.091	0.392	1.000	4.011	>100	>100
0.995	5.298	$g_1(t\|\theta)$	0.411	0.929	0.986	0.995	0.999	1.000	1.000
		$\mathrm{RR}(t\|\theta)$	0.413	0.934	0.991	1.000	1.004	1.005	1.005
		$\mathrm{OR}(t\|\theta)$	0.004	0.066	0.343	1.000	4.921	>100	>100
0.999	6.908	$g_1(t\|\theta)$	0.499	0.968	0.996	0.999	1.000	1.000	1.000
		$\mathrm{RR}(t\|\theta)$	0.499	0.969	0.997	1.000	1.001	1.001	1.001
		$\mathrm{OR}(t\|\theta)$	0.001	0.031	0.250	1.000	7.950	>100	>100

Table 2: Point estimates and standard errors for the parameters of the four proportional-odds models described in Section 5 with the data from the metastatic renal carcinoma trial. The numbers in parenthesis after each model in the first line indicate the equation number in the text.

	Model (28)	Model (29)	Model (30)	Model (31)
$\hat{\eta}_0$	0.584	0.461	2.902	2.600
	(0.124)	(0.131)	(0.592)	(0.814)
$\hat{\eta}_1$		-0.197	0.703	0.547
		(0.0688)	(0.223)	(0.300)
$\hat{\eta}_2$			0.0531	0.0473
			(0.0125)	(0.0176)
$\hat{\theta}_1$	-0.516	-0.469	-0.456	0.312
	(0.168)	(0.169)	(0.171)	(1.194)
$\hat{\theta}_2$				0.388
				(0.460)
$\hat{\theta}_3$				0.0152
				(0.0256)
log-likelihood	-577.8	-573.6	-563.9	-563.3
AIC	1159.6	1153.2	1135.8	1138.6

Standard errors in parentheses. AIC: Akaike Information Criterion.

Table 3: Point estimates and standard errors for the parameters of the four power-probability models described in Section 5 with the data from the metastatic renal carcinoma trial. The numbers in parenthesis after each model in the first line indicate the equation number in the text.

	Model (32)	Model (33)	Model (35)	Model (36)
$\hat{\eta}_0$	0.0267	-0.0467	1.565	1.319
	(0.0774)	(0.0832)	(0.404)	(0.526)
$\hat{\eta}_1$		-0.110	0.463	0.349
		(0.0396)	(0.155)	(0.195)
$\hat{\eta}_2$			0.0358	0.0308
			(0.00882)	(0.0117)
$\hat{\theta}_1$	-0.345	-0.319	-0.298	0.355
	(0.112)	(0.112)	(0.112)	(0.826)
$\hat{\theta}_2$				0.305
				(0.325)
$\hat{\theta}_3$				0.0134
				(0.0182)
log-likelihood	-577.8	-574.1	-564.1	-563.5
AIC	1159.6	1154.2	1136.2	1139.0

Standard errors in parentheses. AIC: Akaike Information Criterion.

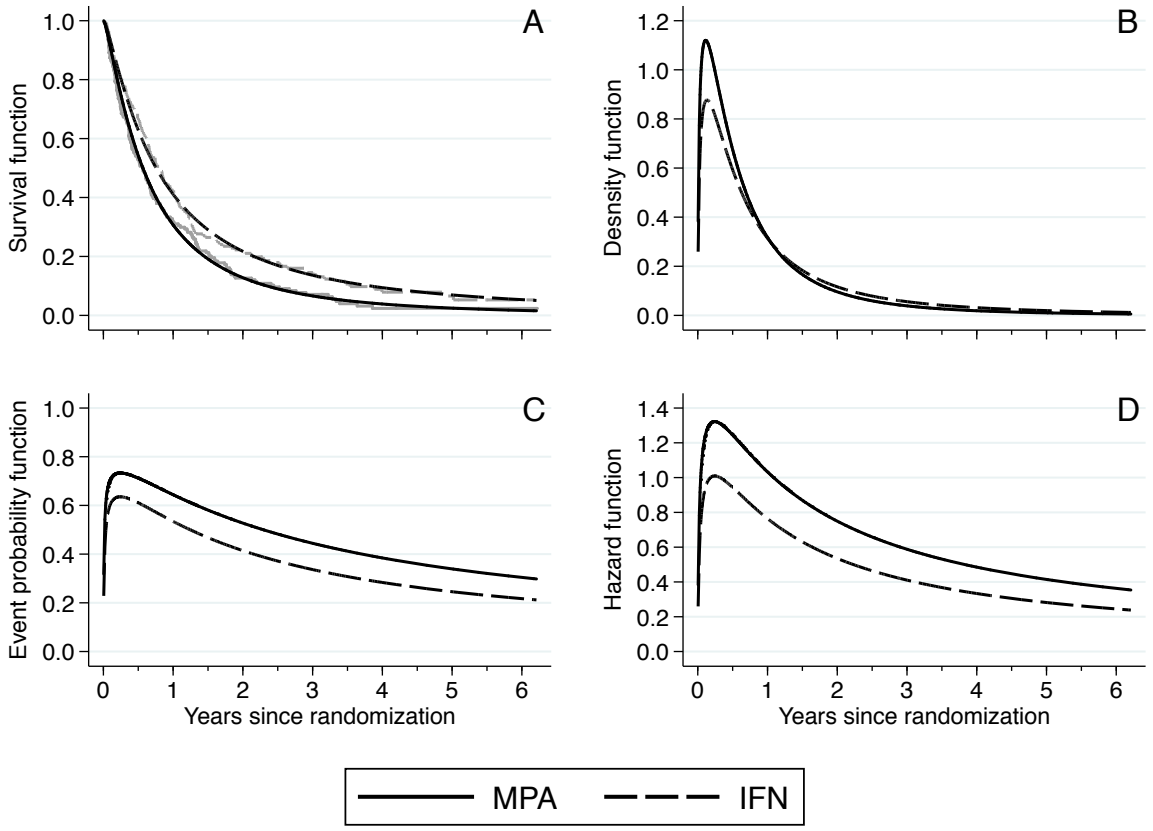

Figure 1: Survival functions (panel A), density functions (B), event-probability functions (C), and hazard functions (D), estimated with the model in equation (30) for each treatment group in the metastatic renal carcinoma trial.

Appendix A

Proof of Corollary 1. First, we show that for all $x > 0$

$$1 - \exp(-x) < x \tag{37}$$

We define the function $z(x) = 1 - \exp(-x) - x$ and its first derivative $z'(x) = \exp(-x) - 1$. Note that $z(0) = 0$ and $z'(x) < 0$ for all $x > 0$. The latter inequality holds because $e(-x) < 1$ for all $x > 0$. Because the functions $z(x)$ and $z'(x)$ are continuous, by the mean value theorem, $z(x) < 0$ for all $x > 0$, which implies inequality (37).

Second, we note that over the support $\mathcal{T}$, the hazard function is continuous and strictly positive, $h(t) > 0$. Replacing x with $h(t)$ in inequality (37) yields $1 - \exp(-h(t)) < h(t)$. By equation (7), the latter inequality implies $g(t) < h(t)$ over the support $\mathcal{T}$.

Proof of Lemma 1. Under the condition stated in Lemma 1 and through equation (12) and the definition of Riemann integrals as limits of Riemann sums, it follows that

$$\begin{aligned}
S(t) &= \exp[-H(t)] \\
&= \exp\left\{\int_0^t \log[\bar{g}(u)]du\right\} \\
&= \exp\left\{\lim_{k\to\infty}\sum_{i=1}^{k}\log[\bar{g}(u_i)]\delta_k\right\} \\
&= \lim_{k\to\infty}\exp\left\{\sum_{i=1}^{k}\log[\bar{g}(u_i)]\delta_k\right\} \\
&= \lim_{k\to\infty}\exp\left\{\sum_{i=1}^{k}\log\left[\bar{g}(u_i)^{\delta_k}\right]\right\} \\
&= \lim_{k\to\infty}\exp\left\{\log\left[\prod_{i=1}^{k}\bar{g}(u_i)^{\delta_k}\right]\right\} \\
&= \lim_{k\to\infty}\prod_{i=1}^{k}\bar{g}(u_i)^{\delta_k}
\end{aligned}$$

Proof of Lemma 2. The first expression follows by integrating by parts the function $h(u) = h_1(u)h_2(u)$, with $h_1(u) = 1$ and $h_2(u) = -\log[\bar{g}(u)]$. The second expression follows by integrating by substitution with $v = \bar{g}(u)$ and $u = \bar{g}^{-1}(v)$ and then by parts.

$$\begin{aligned} H(t) &= \int_0^t h(u)du && (38)\\ &= \int_0^t -\log[\bar{g}(u)]du && (39)\\ &= \int_{\bar{g}(0)}^{\bar{g}(t)} -\log(v)d\ \bar{g}^{-1}(v) && (40)\\ &= \int_{\bar{g}(0)}^{\bar{g}(t)} \bar{g}^{-1}(v)/v\ dv - \left[\log(v)\bar{g}^{-1}(v)\right]\Big|_{\bar{g}(0)}^{\bar{g}(t)} && (41) \end{aligned}$$

The third expression follows by integrating by substitution first and then by parts. With $\bar{g}'$ indicating the first derivative and $\bar{g}^{-1}$ the inverse function,

$$\begin{aligned} H(t) &= \int_0^t h(u)du && (42)\\ &= \int_0^t -\log[\bar{g}(u)]du && (43)\\ &= \int_{\bar{g}(0)}^{\bar{g}(t)} -\frac{\log(v)}{\bar{g}'[\bar{g}^{-1}(v)]}dv && (44) \end{aligned}$$