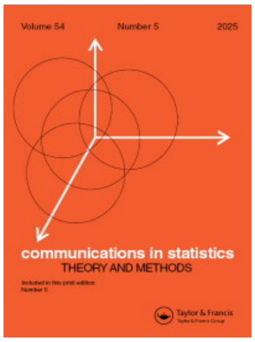

Taylor & Francis
Taylor & Francis Group

Communications in Statistics - Theory and Methods

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/lsta20

New 3-parameter survival distributions from Manly's transform

Rose Baker

To cite this article: Rose Baker (06 Jan 2025): New 3-parameter survival distributions from Manly's transform, Communications in Statistics - Theory and Methods, DOI: 10.1080/03610926.2024.2440004

To link to this article: https://doi.org/10.1080/03610926.2024.2440004

Published online: 06 Jan 2025.

COMMUNICATIONS IN STATISTICS—THEORY AND METHODS
https://doi.org/10.1080/03610926.2024.2440004

New 3-parameter survival distributions from Manly's transform

Rose Baker

School of Business, University of Salford, Salford, UK

ABSTRACT

Two new 3-parameter distributions that generalize the Weibull distribution are introduced. They fit a range of datasets comparably to the generalized gamma and exponentiated Weibull distributions, allowing increasing hazard, decreasing hazard, and bathtub and inverted bathtub hazards. The probability density function can be unimodal, J-shaped, or U-shaped. The survival function is given in closed form, and random numbers can be readily generated. Moments can be evaluated as integrals. For decreasing hazard distributions one distribution can also be a non mixture cure model, a promotion-time model, where the promotion time follows an exponentiated exponential distribution. The second, related distribution does not give a cure model.

ARTICLE HISTORY
8 April 2024
4 December 2024

KEYWORDS
Survival distribution; exponentiated Weibull distribution; generalized gamma distribution; cure model; promotion-time model

1. Introduction

The Weibull distribution has been widely used for the analysis of lifetime data, and can give increasing hazard (increasing force of mortality, or IFOM), decreasing hazard (decreasing force of mortality or DFOM) and constant hazard (exponential) distributions. Some more flexible distributions start by modifying the exponential distribution, e.g., Lemonte (2013), Alotaibi, Nassar, and Elshahhat (2024). However, the Weibull distribution fits increasing and decreasing hazard data so well that it is natural to attempt to accommodate bathtub shaped hazards by generalizing it.

Numerous many 3-parameter generalizations of it have been introduced (see e.g., Cordeiro, Silva, and Nascimento 2020), with the aim of reproducing the U-shaped (bathtub) hazards and modal or upside-down bathtub (UBT) hazards often encountered. These distributions famously include the generalized gamma or Stacy distribution (e.g.Cox, 2007) and the exponentiated Weibull distributions (e.g., Nadarajah, Cordeiro, and Ortega 2013), which usually fit very similarly (Cox, 2014), plus very many others.

The proposed new distributions generalizing the Weibull distribution are based on the exponential transformation of the random variable X to

$$y(x) = \{\exp(\lambda x) - 1\}/\lambda, \tag{1}$$

described by Manly (1976). For small λ, $y(x) = x + \lambda x^2/2 + \cdots$, so that as $\lambda \to 0$, $y \to x$. The function $y(x)$ is convex when $\lambda > 0$, and concave when $\lambda < 0$, so that then $y(x) = (1 - \exp(-|\lambda|x))/|\lambda|$.

CONTACT Rose Baker rose.baker@cantab.net School of Business, University of Salford, Salford, UK.

The inverse transformation is $x = \ln(1 + \lambda y)/\lambda$, and for small λ, $x = y - \lambda y^2/2 + \cdots$. We can call the new distribution with transformed random variable the Y-Weibull distribution (short for 'Yet another extended Weibull-distribution'). It has survival function

$$S(x) = \exp\{-(\frac{\exp(\lambda\alpha x) - 1}{\lambda})^{\beta}\}, \tag{2}$$

where $\alpha > 0$, $\beta > 0$, and λ can be positive or negative. When $\lambda \to 0$ the Weibull survival function $S(x) = \exp(-(\alpha x)^{\beta})$ is regained.

The meaning of the parameters is that α gives the time-scale, β defines the shape of the hazard for early times x, and λ gives the tail behaviour at large x. The survival function is of the form $S(x) = \exp(-H(x))$, where H is a monotonically increasing function, and is thus of the general type described by Gurvich (1997). Although many functional forms for H have been used, the one adopted here has not.

The Y-Weibull distribution can be IFOM, DFOM, BT, or UBT, and is a cure model when $\lambda < 0$ (e.g., Amico and van Keilegom 2018; Peng and Yu 2021). In that case, $S(\infty) = \exp(-1/|\lambda|^{\beta})$. This means that a proportion of events $S(\infty)$ never occur; this is useful in medical decision-making. When $\beta = 1$ and $\lambda > 0$, we have the Gompertz distribution.

Sometimes, especially in medical statistics, a cure model is wanted. Some proportion of patients treated will never relapse from a disease. However, often a cure model is not wanted. For example, in maintenance and reliability, we consider that all devices will fail eventually, and we are often not concerned about the failure mode. Using the Manly transformation in a different way, a related distribution can be derived that never gives a cure model. Here, for $\lambda > 0$, we have the Y-Weibull distribution from (2). However, when $\lambda < 0$, writing for clarity $\eta = |\lambda|$, we invert the transform (1) to take $\alpha x \to \ln(1 + \eta\alpha x)/\eta$. For $\lambda < 0$ we then have the survival function

$$S(x) = \exp\{-(\frac{\ln(1 + \eta\alpha x)}{\eta})^{\beta}\}, \tag{3}$$

which goes to zero as $x \to \infty$. We call this distribution the Z-Weibull. When $\eta \to 0$ it reduces to the Weibull distribution. It can be seen that S is continuous as a function of λ at the join $\lambda = 0$, as is its first derivative with respect to λ. This means that a function minimiser can maximise the log-likelihood function and cross over the join at $\lambda = 0$ with no difficulty. We can think of the Y-Weibull distribution for $\lambda > 0$ and the Z-Weibull distribution for $\lambda < 0$ either as one distribution, or as two distributions joined at the hip for computational convenience.

Next, the remaining properties of the Y-Weibull and Z-Weibull distributions are given, followed by some fits to data. The relation to other distributions is discussed in Appendix A. The text pertains to the Y-Weibull distribution unless otherwise stated.

2. Properties

2.1. Hazard function

From (2) the hazard function is

$$h(x) = \alpha\beta\{\frac{\exp(\lambda\alpha x) - 1}{\lambda}\}^{\beta-1}\exp(\lambda\alpha x).$$

Taking the logarithm of the hazard,

$$\ln h(x) = \ln(\alpha\beta) + (\beta - 1)\ln(\frac{\exp(\lambda\alpha x) - 1}{\lambda}) + \lambda\alpha x. \tag{4}$$

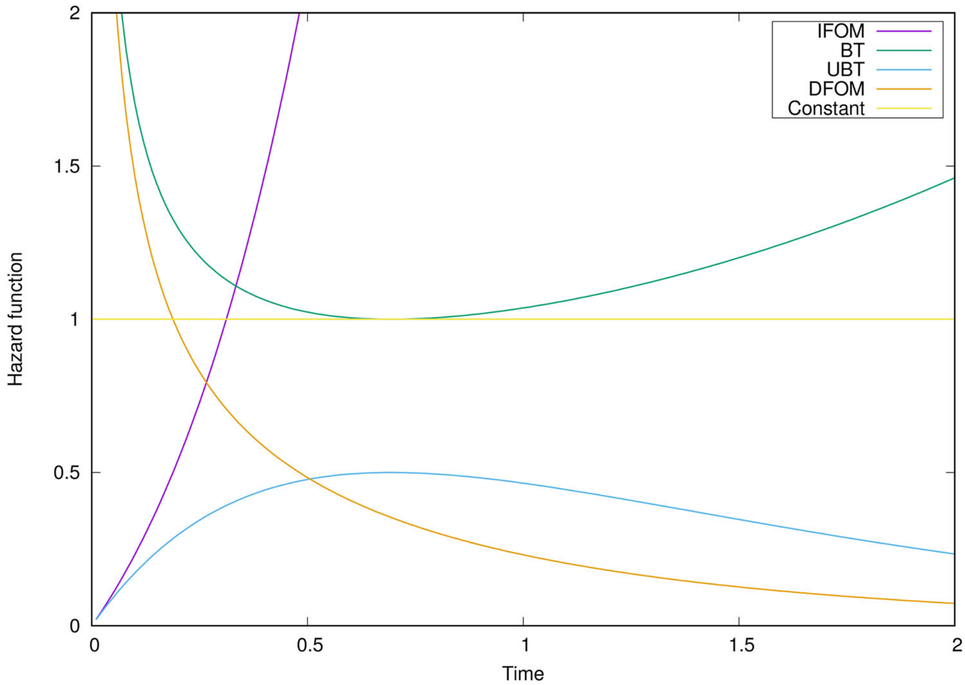

Figure 1. The 5 types of hazard function: increasing (IFOM) (here $\beta = 2, \lambda = 1$), bathtub (BT) (here $\beta = 0.5, \lambda = 1$), upside-down bathtub (UBT) (here $\beta = 2, \lambda = -1$), decreasing (DFOM) (here $\beta = 0.5, \lambda = -1$) and constant ($\lambda = 0, \beta = 1$).

Differentiating, it can be seen that the hazard function has a stationary value x_c when $x_c = \frac{-\ln(\beta)}{\lambda\alpha}$. This occurs at $x_c > 0$ when $\beta > 1, \lambda < 0$ or $\beta < 1, \lambda > 0$. Otherwise, the hazard function increases or decreases monotonically.

Calculating change points for the hazard is much simpler than for the exponentiated Weibull and generalised gamma distributions. These calculations are of practical usefulness, so this is a positive feature. Thus for distributions with bathtub hazard functions, the change point is a marker for the end of the burn-in period. Stationary values of the mean residual life are better measures of the end of the burn-in period because largest expected residual life is exactly what the consumer wants, but these must be computed numerically.

The five hazard types are thus:

1. $\beta > 1, \lambda \geq 0$: increasing force of mortality (IFOM)
2. $\beta > 1, \lambda < 0$: upside-down bathtub (UBT)
3. $\beta < 1, \lambda > 0$: bathtub (BT)
4. $\beta < 1, \lambda \leq 0$: decreasing force of mortality (DFOM)
5. $\beta = 1, \lambda = 0$: constant.

They are illustrated in Figure 1.

For the Z-Weibull distribution when $\lambda < 0$, the hazard function is

$$h(x) = \alpha\beta\{\frac{\ln(1 + \eta\alpha x)}{\eta}\}^{\beta-1}/(1 + \eta\alpha x).$$

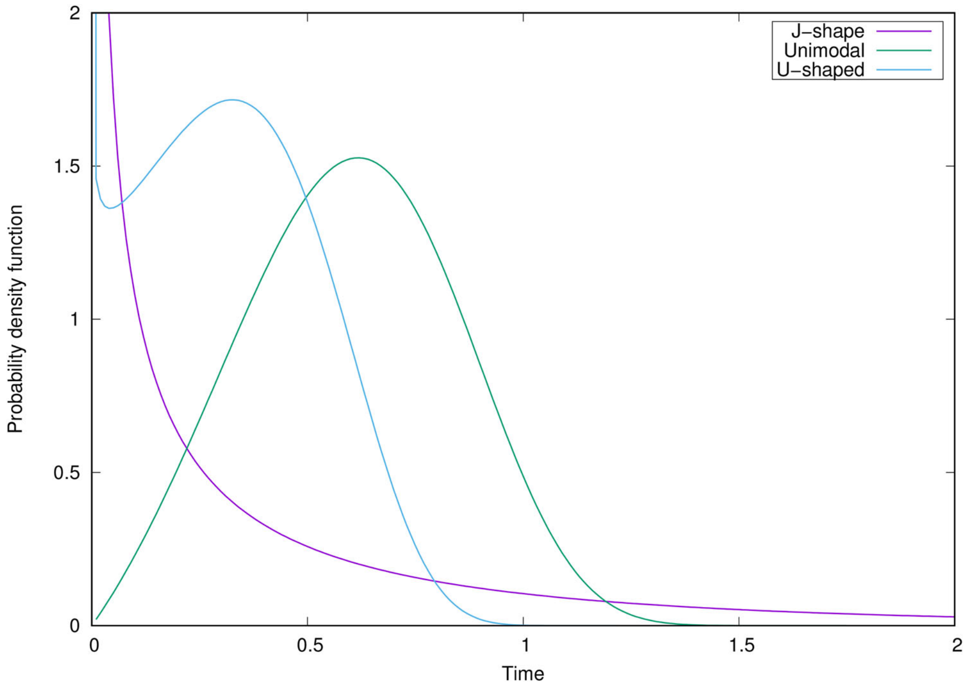

Figure 2. The 3 types of pdf: J-shaped (here $\beta = 0.5, \lambda = -1$), unimodal (here $\beta = 2, \lambda = 1$) and U-shaped (here $\beta = 0.9, \lambda = 4$).

Differentiating this, it can be seen that $h'(x) = 0$ when $x_c = (\exp(\beta - 1) - 1)/\eta\alpha$. For $\beta > 1$, there is a maximum at $x_c > 0$, giving a UBT distribution, and for $\beta \leq 1$ it is DFOM like the Y-Weibull. Similarly it can be shown that if $\beta > 1$ the pdf has a maximum.

2.2. Probability density function (pdf)

The pdf is

$$f(x) = \alpha\beta\{\frac{\exp(\lambda\alpha x) - 1}{\lambda}\}^{\beta-1} \exp(\lambda\alpha x) \exp(-\{\frac{\exp(\lambda\alpha x) - 1}{\lambda}\}^{\beta}).$$

It can be J-shaped, like the exponential distribution, unimodal, or U-shaped, as shown in Figure 2. The U-shape can occur when $\beta < 1$ and λ is large, i.e., there is a bathtub-shaped hazard function and the hazard drops to very low values. Figure 2 shows the 3 types of pdf. The human lifespan has a U-shaped pdf, which this model can approximate, capturing both infant mortality and the increasing mortality due to age. This seems to be the only distribution that can model a U-shaped pdf. Skipping ahead, the Aarset dataset pdf looks U-shaped when histogrammed. The fitted model is also U-shaped, and the fit is much better than the generalised gamma or exponentiated Weibull model fits. Hence the ability to model U-shaped pdfs is sometimes useful in practice.

A proof that the pdf can be unimodal or U-shaped is sketched out: Since $f = hS$, when f is stationary, we have that $\mathrm{d}f/\mathrm{d}x = 0 = (\mathrm{d}h/\mathrm{d}x)S - h^2S$ or $h(x) = \mathrm{d}\ln h(x)/\mathrm{d}x$.

Hence from (4)

$$\alpha\beta\{\frac{\exp(\lambda\alpha x)-1}{\lambda}\}^{\beta-1}\exp(\lambda\alpha x)=\alpha(\beta-1)\lambda\exp(\lambda\alpha x)/(\exp(\lambda\alpha x)-1)+\alpha\lambda.$$

Multiplying through by $\exp(\lambda\alpha x)-1)/\lambda$, rearranging, and adding 1 to each side, we have

$$\mathrm{LHS}(x)\equiv\beta\exp(\lambda\alpha x)=\beta\{\exp(\lambda\alpha x)-1)/\lambda\}^{\beta}\exp(\lambda\alpha x)+1\equiv\mathrm{RHS}(x). \tag{5}$$

From this expression, we can find the modality of the pdf $f(x)$. When $\beta<1,\lambda>0$, then $\mathrm{LHS}(0)=\beta$ and $\mathrm{RHS}(0)=1$, so $\mathrm{LHS}(0)<\mathrm{RHS}(0)$. As $x\to\infty$ also, $\mathrm{LHS}(x)<\mathrm{RHS}(x)$. However, if we keep $\lambda\alpha x$ constant and increase λ, the left-hand side stays constant while the right-hand side decreases to 1, so that for large enough λ and some x, $\mathrm{LHS}(x)>\mathrm{RHS}(x)$. The pdf thus has two stationary values, corresponding to the U-shape in Figure 2. The modes can be found numerically by solving (5).

Similarly, when $\beta>1$, $\mathrm{LHS}(0)=\beta$, so $\mathrm{LHS}(0)>\mathrm{RHS}(0)$. For large x, however, $\mathrm{LHS}(x)<\mathrm{RHS}(x)$. Thus, there is one mode, the unimodal case in Figure 2.

2.3. Random numbers

Since S is computable, given a uniformly-distributed random number U, we have that

$$X=\frac{\ln\{\lambda(-\ln(U))^{1/\beta})+1\}}{\lambda\alpha} \tag{6}$$

is a random number from the Y-Weibull distribution. However, if $\lambda<0$, there is a probability that events will not occur. Thus, given a time cutoff T, if $U<\exp(-1/|\lambda|^{\beta})$ we take the event as censored at T.

For the Z-Weibull distribution, when $\lambda<0$, random numbers can be generated as

$$X=\frac{\exp(\eta(-\ln(U))^{1/\beta}-1}{\eta\alpha}.$$

2.4. Moments

The median is derivable as

$$x_m=\frac{\ln\{\lambda\ln(2)^{1/\beta}+1\}}{\lambda\alpha}.$$

It exists if $\lambda>(-\ln(2))^{-1/\beta}$. The moments cannot be found in closed form, but for example the mean μ is given by

$$\mu=\int_0^{\infty}\frac{\exp(-y^{\beta})\,\mathrm{d}y}{\alpha(1+\lambda y)}.$$

This follows from using (1) and the result that $\mu=\int_0^{\infty}S(u)\,\mathrm{d}u$. When $\lambda<0$ the mean is infinite.

For the Z-Weibull the median is $\frac{\exp(\eta 2^{1/\beta})-1}{\eta\alpha}$. The mean is $\mu=\alpha^{-1}\int_0^{\infty}\exp(-y^{\beta}+\eta y)\,\mathrm{d}y$. It exists if $\beta>1$ or $\beta=1$ and $\eta<1$, otherwise it is infinite.

2.5. Cure model

When $\lambda < 0$, we can write

$$S(x) = \exp(-\{\frac{1-\exp(-|\lambda|\alpha x)}{|\lambda|}\}^{\beta}), \tag{7}$$

This is a non mixture cure model of the promotion-time or 'first activation scheme' type $S = \exp(-\gamma_0 F(x))$, where F is a distribution function. Here we have that $F(x) = (1-\exp(-|\lambda|\alpha x))^{\beta}$, an exponentiated exponential distribution, while $\gamma_0 = |\lambda|^{-\beta}$. Note that γ_0 and F seem to have the same parameters, but are in fact independent, because the factor of α is arbitrary; we could take $F(x) = (1-\exp(-\phi x))^{\beta}$, $\gamma_0 = (\alpha/\phi)^{\beta}$, so the new parameters could be ϕ, β, γ_0.

The error and confidence intervals on the cured proportion can be found from the covariance matrix of fitted model parameters. Since β and $|\lambda|$ are required, an accurate method is to generate random numbers from the bivariate normal distribution with the fitted covariance matrix, and so compute the standard error and confidence interval for $S(\infty)$.

To test whether the cured proportion is positive, one needs to test that $\lambda < 0$. This is easily done either using the fitted standard error on λ in a Wald test, or maximising the likelihood again setting $\lambda = 0$ (a Weibull fit) and doing a chi-squared test. Note that because λ can take either sign, we do not have the inferential problem caused by the cure proportion being at the end of its range that is faced by mixture models.

2.6. Inclusion of covariates

Covariates z_i can be included by regressing $\alpha = \exp(\sum_{i=1}^{m} \beta_i z_i)$ as one does for the Weibull distribution. This is an AFT (accelerated failure time) model for the latency part of the cure model. The probability of cure $S(\infty)$ can also be regressed on covariates, to allow the incidence of cure to be a function of covariates. This can be achieved by a linear regression on covariates for $\ln(-\lambda)$. For the Z-Weibull distribution, regression of shape parameters β, λ on covariates would not usually be done.

3. Inference

3.1. Simulations

Fitting the datasets in Table 1, the median values of β, λ were found for the four non trivial hazard types shown in Figure 1, for a sample size of 128. This is the size of the largest dataset fitted. Simulated datasets were created using (6) and setting $\alpha = 1$, and fitted by maximum-likelihood. In the simulations, parameters were fitted starting from the true values.

Table 2 shows the bias, standard error, and standard error of the mean for 10,000 simulations from each of these four groups, and for the exponential distribution. The small standard error of the mean confirms that these estimates of bias are sufficiently accurate.

Bias is taken as e.g.$\hat{\beta} - \beta$, and is small, except when β is large, and especially so when λ is also large. In this case, the tail cuts off quickly, and there are few events to enable a good determination of λ. Hence in these cases, the bias is large and the standard error of $\hat{\lambda}$ very large. The percentages of simulated datasets correctly classified in the first 4 groups are 100, 84.5, 100, and 98.8. The parameter λ is easily estimated with the wrong sign when small and negative (giving a longer tail) and when β is large.

Table 1. Details of the datasets fitted and where to find them.

Dataset	Sample size	Description	Source
Bebb3	60	Elec. appliance	Bebbington, Lai, and Zitikis (2006) tab. 3
Bebb4	36	Generators	Bebbington, Lai, and Zitikis (2006) tab. 4
Bebb5	128	Load-haul dump trucks	Bebbington, Lai, and Zitikis (2006) tab. 5
Bebb6	14	Reactor pumps	Bebbington, Lai, and Zitikis (2006) tab. 6
Flood	72	Wheaton river exceedences	Akinsete, Famoye, and Lee (2008)
Bladder	128	Bladder cancer	Shanker et al. (2016)
Leuk	51	Remission to relapse for leukaemia	Xie, Tang, and Goh (2002)
Aarset	50	Devices	Xie, Tang, and Goh (2002)
Repair	40	Transceiver repair time	Xie, Tang, and Goh (2002)
Fibres	63	Strength of 1.5 cm glass fibres	Shanker et al. (2016)
Pigs	72	Survival of guinea pigs	Shanker et al. (2016)
Glass	31	Strength of aeroplane windows	Shanker et al. (2016)
Bank	100	Bank waiting times	Ghitany, Atieh, and Nadarajah (2008)
Ballbear	23	Ball-bearing lifetimes	Xie, Tang, and Goh (2002)
Guinea	65	Guinea-pig survival	Xie, Tang, and Goh (2002)
Allogen	45	Allogenic transplant for leukaemia	Xie, Tang, and Goh (2002)
Autolog	45	Autologous transplant for leukaemia	Xie, Tang, and Goh (2002)
Coupon	101	Aluminium coupons	Xie, Tang, and Goh (2002)
Pressure	20	Pressure vessels	Xie, Tang, and Goh (2002)
Fluid	19	Breakdown of insulating fluid	Xie, Tang, and Goh (2002)

The event is always failure for devices unless otherwise stated.

Also, in Table 3 fits are shown for one simulation. From the fits, it can be seen that in general the departure of the fitted parameter from the true value is comparable with the computed standard error.

3.2. Fits to data

Fitting can be done by a likelihood-based method such as likelihood maximisation. Because we know the survival distribution, any type of censoring is readily coped with. Derivatives of the log-likelihood are not hard to compute, but one can maximise the log-likelihood quickly enough using a quasi-Newton method that does not require derivatives. Fitting is best started from an exponential distribution with the sample mean, so that $\alpha = 1/\bar{x}, \beta = 1, \lambda = 0$.

Careful coding will test for the case that $|\lambda| < \epsilon$, some small number such as 10^{-4}, and then default to the Weibull distribution. Better, one can approximate $\exp(\lambda\alpha x) - 1)/\lambda \simeq x + \lambda x/2 + \lambda^2 x^2/6$ for small $|\lambda|$.

Table 1 shows details of 20 datasets used to assess the model. All are available in the public domain. The source given here is not always the original publication.

Table 4 shows the Akaike Information Criterion (AIC) of fits to these datasets. Note that all models have 3 parameters, so effectively lower AIC means higher likelihood. The cure percentage is also shown; this is usually zero and of course can only be non zero for the Y-Weibull distribution and the Xie, Tang, and Goh (2002) distribution, and for UBT hazards. Twice here the hazard extremum occurs beyond the maximum time in the data (glass and coupon datasets), and then BT is taken as DFOM and UBT as IFOM. These classifications are taken from the Y-Weibull distribution fit.

It can be seen that the new distributions perform comparably with the generalised gamma and exponentiated Weibull distributions, sometimes fitting best for all types, such as IFOM, BT, and UBT. The exponential transformation means that they can accommodate exponen-

Table 2. Results of fitting 10000 simulated datasets, showing biases of parameter estimates, standard errors of estimates, and the standard error of the computed bias.

Parameter	True value	Bias	s.e.	s.e.m
α	1	0.0268	0.1948	0.0019
β	0.6	0.0095	0.0583	0.0006
λ	-0.3	-0.0068	0.0614	0.0006
α	1	-0.0317	0.1442	0.0014
β	2.5	-0.0192	0.4055	0.0041
λ	-0.3	0.0931	0.2665	0.0027
α	1.	-0.0065	0.2401	0.0024
β	0.6	0.0024	0.0727	0.0007
λ	0.8	0.1861	0.6278	0.0063
α	1.	-0.0416	0.2657	0.0027
β	2.5	0.0050	0.5639	0.0056
λ	0.8	0.7630	3.2386	0.0324
α	1	-0.0246	0.1448	0.0014
β	1	-0.0066	0.1133	0.0011
λ	0	0.0732	0.2002	0.0020

Table 3. Results of fitting simulated datasets with $\alpha = 1$, showing fitted values and standard errors.

β	λ	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\lambda}$
0.6	-0.3	1.33 (0.19)	0.607 (0.044)	-0.25 (0.042)
2.5	-0.3	0.98 (0.14)	2.34 (0.34)	-0.167 (0.22)
0.6	0.8	1.22 (0.24)	0.59 (0.06)	0.74 (0.34)
2.5	0.8	0.94 (0.29)	2.30 (0.51)	1.19 (1.17)
1	0	0.98 (0.17)	0.94 (0.12)	0.19 (0.24)

Table 4. AIC for model fits of the Y-Weibull model compared to two popular 3-parameter distributions, the generalised gamma and the exponentiated Weibull, and with the modified Weibull of Xie, Tang, and Goh (2002), extended to allow $\lambda < 0$.

Dataset	Type	Cure %	Y-Weibull	Gen. gamma	Exp. Weibull	Xie, Tang, and Goh (2002)
Bebb3	BT	0	216.78	216.22*	216.32	216.43
Bebb4	BT	0	142.24*	142.84	142.81	142.53
Bebb5	DFOM	0	364.29	362.63*	362.66	363.93
Bebb6	BT	0	30.14*	31.22	31.06	31.20
Flood	BT	0	507.32*	508.13	508.05	507.56
Bladder	UBT	0.44	826.84*	827.71	827.36	827.61
Leuk	UBT	1.04	677.74	676.97*	677.18	678.18
Aarset	BT	0	457.23*	470.50	467.99	460.29
Repair	UBT	1.38	193.16	191.44	189.45*	193.88
Fibres	IFOM	0	34.81*	43.06	35.35	36.38
Pigs	UBT	0.05	193.88*	194.46	194.18	194.98
Glass	IFOM	0.06	216.65	216.09	215.55*	217.11
Bank	IFOM	0	640.94	639.91*	639.94	641.69
Ballbear	UBT	0.008	232.43	232.00	231.95*	232.96
Guinea	IFOM	0	876.56	877.49	871.21*	876.20
Allogen	UBT	27.0	474.73*	480.33	474.91	474.82
Autolog	UBT	23.0	391.18*	426.05	407.65	398.71
Coupon	IFOM	0	1493.99*	1495.79	1497.92	1494.37
Pressure	BT	0	291.66*	292.21	292.02	293.14
Fluid	BT	0	212.41*	213.75	213.18	213.95

Cure percentage and distribution type are shown, and the best fitting model marked with an asterisk. Only the allogen and autolog datasets are censored.

tially increasing hazards. The presence of only one DFOM distribution in the datasets may be because some distributions regarded as DFOM in fact have a hazard function that increases at large times, making them BT. This seems not unreasonable. As expected, the generalised

Table 5. Kolmogorov-Smirnov *D* statistic and p-value in parentheses for model fits of the Y-Weibull model compared to two popular 3-parameter distributions, the generalised gamma and the exponentiated Weibull, and with the modified Weibull of Xie, Tang, and Goh (2002), extended to allow $\lambda < 0$.

Dataset	Y-Weibull	Gen. gamma	Exp. Weibull	Xie, Tang, and Goh (2002)
Bebb3	0.0548 (0.992)	0.086 (0.714)	0.0547 (0.912)	0.0522 (0.996)*
Bebb4	0.079 (0.854)*	0.1020 (0.825)	0.1024 (0.824)	0.1046 (0.804)
Bebb5	0.0574 (0.781)	0.0428 (0.970)	0.0413 (0.979)*	0.0536 (0.846)
Bebb6	0.157 (0.850)	0.190 (0.645)	0.129 (0.962)*	0.056 (0.846)
Flood	0.0994 (0.455)*	0.1082 (0.350)	0.1074 (0.358)	0.1037 (0.401)
Bladder	0.0416 (0.977)	0.0466 (0.618)	0.040 (0.995)*	0.046 (0.944)
Leuk	0.075 (0.927)	0.0674 (0.967)	0.0659 (0.976)*	0.0821 (0.867)
Aarset	0.125 (0.391)*	0.297 (0.0002)	0.1872 (0.052)	0.1392 (0.266)
Repair	0.125 (0.431)	0.136 (0.424)	0.126 (0.520*	0.1262 (0.519)
Fibres	0.1205 (0.300)*	0.194 (0.424)	0.1304 (0.218)	0.136 (0.1785)
Pigs	0.1196 (0.238)	0.1139 (0.291)*	0.1146 (0.283)	0.1192 (0.242)
Glass	0.1558 (0.408)	0.1552 (0.412)	0.1493 (0.462)*	0.1636 (0.348)
Bank	0.0439 (0.989)	0.0359 (0.999)*	0.0362 (0.999)	0.0471 (0.977)
Ballbear	0.1219 (0.862)	0.1187 (0.992)	0.1080 (0.939)*	0.1367 (0.752)
Guinea	0.1294 (0.207)	0.1184 (0.303)*	0.1255 (0.241)	0.1353 (0.171)
Coupon	0.048 (0.970)	0.0377 (0.998)*	0.0478 (0.972)	0.0539 (0.924)
Pressure	0.0836 (0.998)	0.1152 (0.940)	0.0645 (0.999)*	0.0766 (0.999)
Fluid	0.1084 (0.971)*	0.2432 (0.181)	0.1343 (0.875)	0.1378 (0.836)

The best fitting model is marked with an asterisk.

gamma and exponentiated Weibull perform very similarly. The Y-Weibull usually performs better than the modified Weibull of Xie, Tang, and Goh (2002), being better in 17 cases and worse in 3.

Table 5 shows the results of computing the Kolmogorov-Smirnov goodness of fit test for the 18 datasets without censoring. The formulae given in Press et al. (2007) were used for this. The p-values will be over-optimistic, as no allowance is made for estimating the parameters, but the results still give the relative goodness of fit, and the adequacy of the fitted curve to approximate the data. The Y-Weibull is now best 6 times, the generalized gamma best 4 times, the exponentiated Weibull best 8 times, and the modified Xie distribution best once. There is usually little difference in the fit, the exception being the Aarset dataset, where the generalized gamma and exponentiated Weibull perform poorly.

The Z-Weibull distribution is identical to the Y-Weibull when $\lambda > 0$, but when $\lambda < 0$ it performs differently. Thus for the bladder cancer dataset, the AIC was 825.97, slightly better than for the Y-Weibull. However, for the 'allogen' dataset, where there is a large cure fraction, the AIC was 484.50, substantially worse than for the Y-Weibull. Fitting this distribution in addition to the Y-Weibull can indicate whether there is need for a cure model or not. Can a flexible non cure model fit the data as well, or better?

The conclusion is that the new models can fit a broad sample of datasets comparably with current models. Among these, the exponentiated Weibull model performed better than the generalized gamma model.

3.3. Testing

There is a huge number of ageing classes of survival distributions (e.g., Lai and Xie 2006), and very many tests that could be done. Focussing on tests for λ, a test that $\lambda > 0$ gives a test for BT hazard-shape vs DFOM that might be applied when when $\hat{\beta} < 1$, and a test that $\lambda < 0$ is a test for UBT vs IFOM applied when $\hat{\beta} > 1$; this logic applies to both Y-Weibull and Z-Weibull distributions. More usefully perhaps, a test that $\lambda < 0$ is a test for a positive cure fraction for the Y-Weibull distribution.

These tests are best done by also fitting the Weibull model, when twice the increase in log-likelihood on moving to the Y-Weibull model is distributed as chi-squared with one degree of freedom. Thus for the bladder cancer dataset, the corresponding signed normal score is $z = -2.71$ (λ negative). From the (less reliable) Wald test, we would have $z = -3.1$, in fair agreement. A score test based on the Weibull distribution fit was found to have low power, and is not described further.

For the Y-Weibull distribution, negative λ leads to a cured proportion, but this may be a statistical artifact, because negative λ is a sign of a strongly DFOM or UBT-shaped distribution, as well as inevitably resulting in a cured percentage. Thus for the bladder cancer dataset, despite the test for negative λ being strongly significant, the 95% confidence interval for the cured percentage was $(0, 2.2\%)$, consistent with zero.

As mentioned, the Z-Weibull distribution, which has zero cured percentage, fitted better. This encourages the belief that the apparent cured fraction is caused by a hazard function that is strongly decreasing, but not necessarily fast enough to give a cured fraction. As indicated in Peng and Yu (2021), deciding whether there is a cured percentage is a vexed question, and more background knowledge than is given by fitting a statistical model is needed to come to a conclusion.

4. Conclusions

The new distributions accommodate all hazard types, fit on average as well as any 3-parameter survival distribution, and the Y-Weibull can also function as a cure model. Thus, this distribution could be useful in modelling both reliability and medical data. It is straightforward from the fit to data to determine which of the four hazard-function classes the distribution belongs to. This will aid decision-making: for devices with bathtub-shaped hazard functions one must then determine the optimum burn-in period, and with increasing hazards there will be an optimum lifetime before preventative replacement.

In the medical case, it offers a simple test of whether the cured proportion is non zero, important for subsequent medical management of a condition such as cancer.

A limitation of this study is that one can never have too much experience in fitting models, but 20 datasets are sufficient to give an indication of the model's usefulness. The bivariate case arises occasionally, when there are two measures of usage, and this has not been considered, beyond recommending using a copula.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

Akinsete, A., F. Famoye, and C. Lee. 2008. The beta-Pareto distribution. *Statistics* 42 (6):547–63. doi: 10.1080/02331880801983876.

Alotaibi, R., M. Nassar, and A. Elshahhat. 2024. A new extended Pham distribution for modelling cancer data. *Journal of Radiation Research and Applied Sciences* 17 (3):100961. doi: 10.1016/j.jrras.2024.100961.

Amico, M., and I. Van Keilegom. 2018. Cure models in survival analysis. *Annual Review of Statistics and Its Application* 5 (1):311–42. doi: 10.1146/annurev-statistics-031017-100101.

Bebbington, M., C.-D. Lai, and R. Zitikis. 2006. useful periods for lifetime distributions with bathtub shaped hazard rate functions. *IEEE Transactions on Reliability* 55 (2):245–51. doi: 10.1109/TR.2001.874943.

Cordeiro, G. M., R. B. Silva, and D. C. Nascimento. 2020. *Recent advances in lifetime and reliability models*. Sharjah: Bentham Books.

Ghitany, M. E., B. Atieh, and S. Nadarajah. 2008. Lindley distribution and its application. *Mathematics and Computers in Simulation* 78 (4):493–506. doi: 10.1016/j.matcom.2007.06.007.

Gurvich, M. R., A. T. Dibenedetto, and S. V. Ranade. 1997. A new statistical distribution for characterizingthe random strength of brittle materials. *Journal of Materials Science* 32 (10):2559–64. doi: 10.1023/A:1018594215963.

Lai, C.-D., and M. Xie. 2006. *Stochastic ageing and dependence for reliability*. Berlin: Springer.

Lemonte, A. J. 2013. A new exponential-type distribution with constant, decreasing, increasing, upside-down bathtub and bathtub-shaped failure rate function. *Computational Statistics & Data Analysis* 62:149–70. doi: 10.1016/j.csda.2013.01.011.

Manly, B. F. J. 1976. Exponential data transformations. *The Statistician* 25 (1):37. doi: 10.2307/2988129.

Nadarajah, S., G. M. Cordeiro, and E. M. M. Ortega. 2013. The exponentiated Weibull distribution: A survey. *Statistical Papers* 54 (3):839–77. doi: 10.1007/s00362-012-0466-x.

Peng, Y., and B. Yu, 2021. *Cure models: Methods, applications and implementation*. London: CRC Press.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. 2007. *Numerical Recipes: the art of scientific computing*. Cambridge: Cambridge University Press.

Shanker, R., K. K. Shukla, R. Shanker, and T. A. Leonida. 2016. On modeling of lifetime data using two-parameter gamma and Weibull distributions. *Biometrics & Biostatistics International Journal* 4 (5):1–6.

Xie, M., Y. Tang, and T. N. Goh. 2002. A modified Weibull extension with bathtub-shaped failure rate function. *Reliability Engineering & System Safety* 76 (3):279–85. doi: 10.1016/S0951-8320(02)00022-4.

Appendix A: Related distributions

As mentioned, the Y-Weibull distribution generalises the Weibull when $\lambda \neq 0$, and the Gompertz when $\beta \neq 1$. The model (2) is also similar to the model

$$S(x) = \exp(-(\exp(\lambda(\alpha x)^{\beta}) - 1)/\lambda)$$

(our parameterisation) proposed by Xie *et al* (2002). Here αx is raised to a power and then the exponential transformation is applied, whereas in (2) αx is transformed and then raised to a power. The Xie, Tang, and Goh (2002) model can also produce a cure model when $\lambda < 0$, when the embedded distribution is a Weibull distribution. Interestingly, Xie, Tang, and Goh (2002) only consider the case $\lambda > 0$, and do not seem to realise that their distribution can also model the UBT case and yield a cure model. It will be shown that this distribution does not in general fit quite as well as the Y-Weibull. As a cure model, the proportion cured is simpler, being $S(\infty) = \exp(-1/|\lambda|)$.

The Z-Weibull distribution is a generalization of the Pareto or Lomax distribution, to which it reduces when $\beta = 1$.

The distribution of the earliest of n random variables from a survival distribution has survival function $S(x)^n$. This still has the Y-Weibull distribution. The same invariance holds for the Xie *et al* distribution, but not for the generalised gamma or exponentiated Weibull.

Three parameters is quite enough for most datasets encountered, but a possibility for large datasets would be to raise x in (2) to a power, so obtaining the distribution

$$S(x) = \exp(-\{\frac{\exp(\lambda(\alpha x)^{\gamma}) - 1}{\lambda}\}^{\beta}), \tag{A.1}$$

where $\gamma > 0$. This distribution would then give the exponentiated Weibull distribution as the embedded distribution function $F(x)$ in the cure model.