Pitfalls of proportional hazards regression methods to assess sex-specific racial differences in survival

in Eaton WW, Roth KB,…, Muñoz A. The Relationship of Mental and Behavioral Disorders to All-cause Mortality in a 27-year Follow-up of Four Epidemiologic Catchment Area Samples. Am J Epidemiol 2013

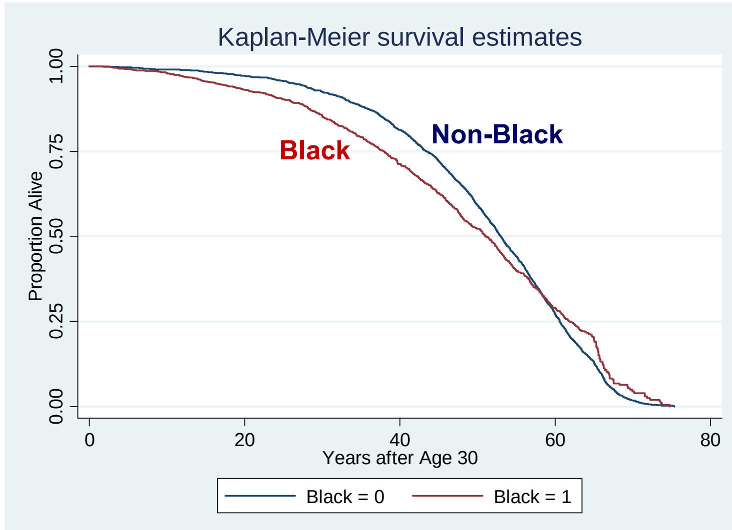
Alvaro Muñoz, PhD
Professor of Epidemiology

Survival after age 30 among Females by race

	Black		non-Black	
	deceased	alive	deceased	alive
N	841	1,805	2,950	3,358
Age at entry	64 [50, 72]	33 [30, 46]	70 [64, 77]	35 [30, 52]
Age at exit	76 [66, 85]	59 [52, 72]	83 [76, 90]	62 [53, 78]
P-Y on study	11,114	42,114	36,984	80,809

Survival after age 30 among Females by race

sts graph if Female==1, by(Black) xtitle(Years after Age 30) ytitle(Proportion Alive)



Standard Cox Regression

$$h(t; black) = h(t; non-black)e^{\beta}$$

stcox Black if Female==1

```
No. of subjects =        8953                Number of obs   =      8953
No. of failures =       3791
Time at risk    =  171023.1643
                                             LR chi2(1)      =      1.41
Log likelihood  =  -28178.581                Prob > chi2     =    0.2351

------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       Black |  .9545171    .037576    -1.18   0.237     .8836389     1.03108
------------------------------------------------------------------------------
```

e^{β}

Poor summary: no difference by race and if anything, black race is protective

Cox as disguised Weibull

$$T \sim Weibull(\alpha + \beta\, black, \sigma)$$

```
streg Black if Female==1, dist(weibull)

Weibull regression -- log relative-hazard form

No. of subjects =        8953                    Number of obs   =      8953
No. of failures =        3791
Time at risk    =  171023.1643
                                                  LR chi2(1)     =      1.06
Log likelihood  =   -909.33593                   Prob > chi2     =    0.3042

------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       Black |  .9606234   .0376984    -1.02   0.306     .889506    1.037427
------------------------------------------------------------------------------
```

$e^{-\beta/\sigma}$

Poor summary: inferences identical to Cox's

Cox Regression allowing for time dependency

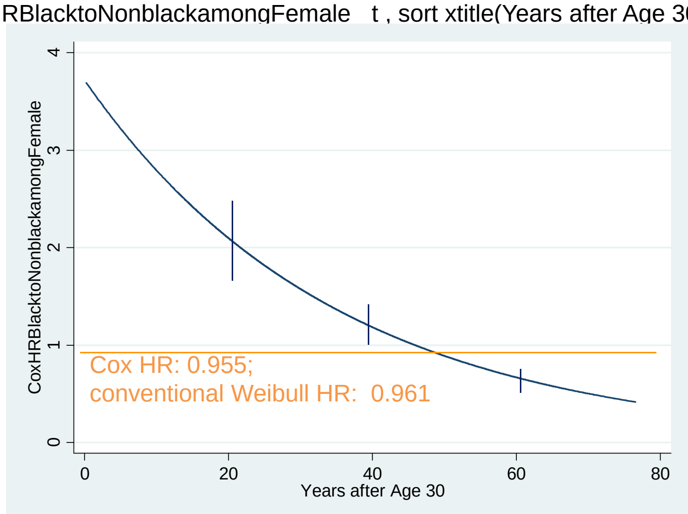
$$h(t; black) = h(t; non-black)e^{\beta+\gamma t}$$

```
stcox Black if Female==1, tvc(Black) texp( _t ) nohr

No. of subjects =        8953                    Number of obs   =      8953
No. of failures =        3791
Time at risk    =  171023.1643
                                                 LR chi2(2)      =     98.78
Log likelihood  =   -28129.896                   Prob > chi2     =    0.0000

------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
main         |   β=
       Black |    1.31299   .1446371    9.08   0.000     1.029506    1.596473
-------------+----------------------------------------------------------------
tvc          |   γ=
       Black |  -.0286116   .0029663   -9.65   0.000    -.0344254   -.0227978
------------------------------------------------------------------------------
```

Significant downward trend of HR

Hazard Ratio from a Cox Model allowing for time dependency

gen CoxHRBlacktoNonblackamongFemale= exp([main]Black + [tvc]Black * _t)
line CoxHRBlacktoNonblackamongFemale _t , sort xtitle(Years after Age 30)



Parametric Models: quantiles in addition to hazards

British MDs: Survival From Age 35, Continued

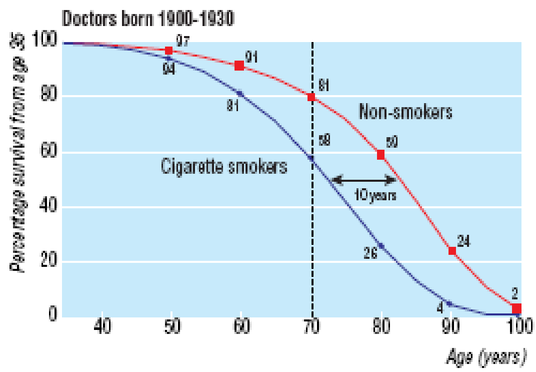


Fig 3 Survival from age 35 for continuing cigarette smokers and lifelong non-smokers among UK male doctors born 1900-1930, with percentages alive at each decade of age

Doll et al., BMJ, 2004doi:10.1136/bmj.38142.554479.AE

Parametric Models

GG conventional $T \sim GG(\alpha + \beta\, \boldsymbol{black}, \sigma, \kappa)$

times are proportional: exp(β)

streg Black if Female==1, dist(gamma) time /*to display beta, sigma, kappa */

	_t	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
β	Black	.0310971	.0069531	4.47	0.000	.0174692	.044725
α	_cons	4.119847	.0062218	662.17	0.000	4.107652	4.132041
	/ln_sig	-1.914397	.0348651	-54.91	0.000	-1.982731	-1.846063
	/kappa	2.499254	.1083305	23.07	0.000	2.28693	2.711578
	sigma	.1474307	.0051402			.1376927	.1578575

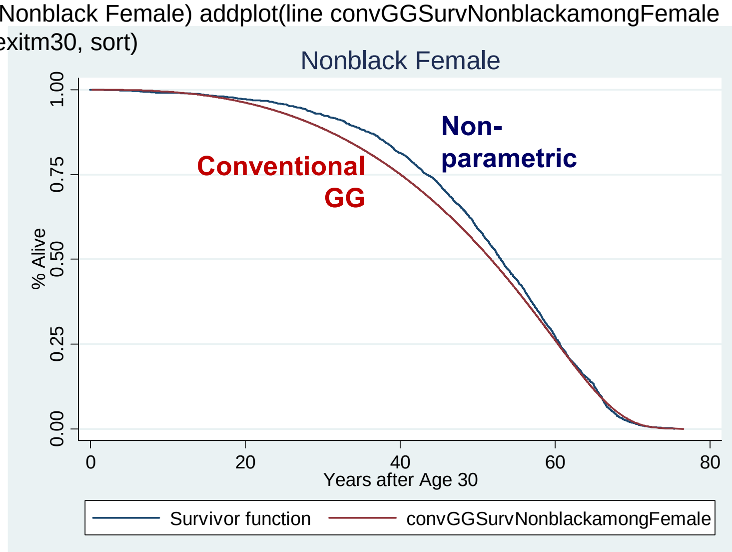
exp(0.031) = 1.03

Poor summary: Blacks live 3% longer

```
*For Non-black Females
gen convGGSurvNonblackamongFemale=
      gammaptail(e(kappa)^(-2),(e(kappa)^(-2))*(exp(_t]_cons)*ageexitm30)^(e(kappa)/e(sigma)))
replace convGGSurvNonblackamongFemale= 1-convGGSurvNonblackamongFemale if e(kappa)<0
```

Fit of conventional GG to non-parametric survival for Non-black Females

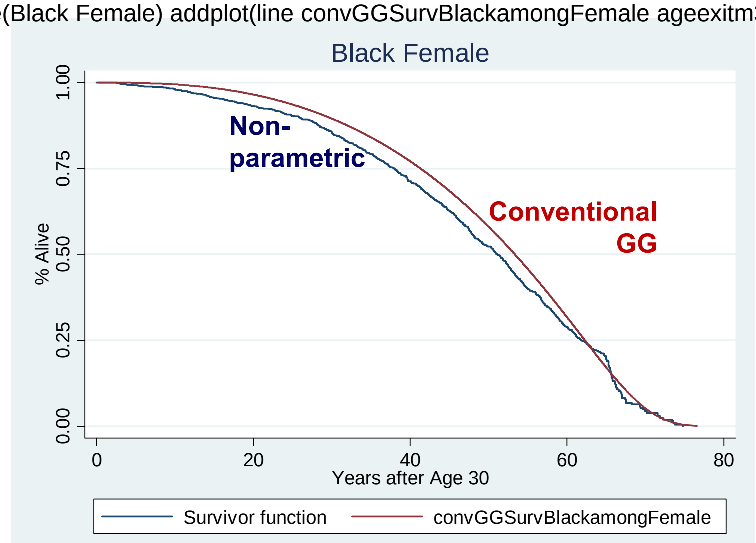
sts graph if Female==1 & Black==0, surv xtitle(Years after Age 30) ytitle(% Alive) /// title(Nonblack Female) addplot(line convGGSurvNonblackamongFemale ageexitm30, sort)



Poor fit for Non-blacks

Fit of conventional GG to non-parametric survival for Black Females

sts graph if Female==1 & Black==1, surv xtitle(Years after Age 30) ytitle(% Alive) /// title(Black Female) addplot(line convGGSurvBlackamongFemale ageexitm30, sort)



and Poor fit for Blacks too

GG non-conventional (saturated)

$$T \sim GG(\alpha + \beta\, black, exp(\gamma_0 + \gamma_1 black), \kappa_0 + \kappa_1 black)$$

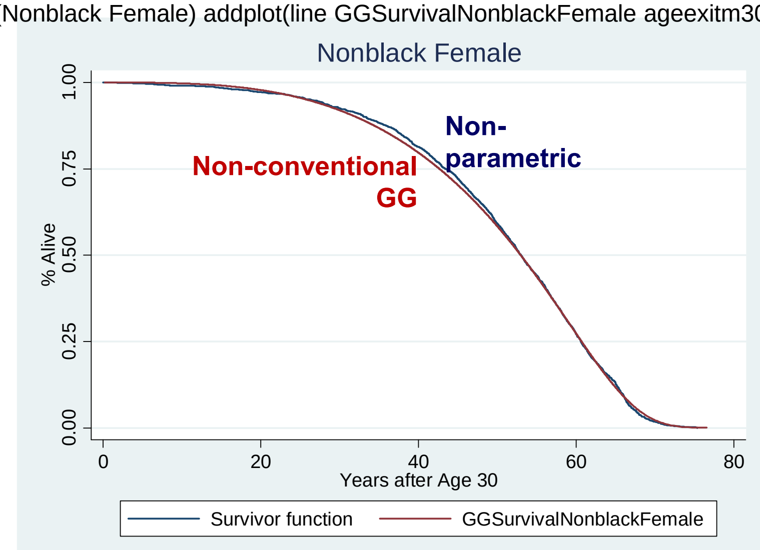
streg Black if Female==1, anc(Black) anc2(Black) dist(gamma)

	_t	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
	_t						
β	Black	.0786024	.0242907	3.24	0.001	.0309936	.1262113
α	_cons	4.107777	.006334	648.53	0.000	4.095363	4.120192
	ln_sig						
γ_1	Black	-.0521234	.1767268	-0.29	0.768	-.3985016	.2942549
γ_0	_cons	-1.918588	.0342164	-56.07	0.000	-1.985651	-1.851525
	kappa						
κ_1	Black	1.266314	.6381863	1.98	0.047	.0154921	2.517136
κ_0	_cons	2.111767	.10096	20.92	0.000	1.913889	2.309645

```
*For Non-black Females
gen convGGSurvNonblackamongFemale=
     gammaptail(e(kappa)^(-2),(e(kappa)^(-2))*(exp(_t]_cons)*ageexitm30)^(e(kappa)/e(sigma)))
replace convGGSurvNonblackamongFemale= 1-convGGSurvNonblackamongFemale if e(kappa)<0
```

Fit of non-conventional GG to non-parametric survival for Non-black Females

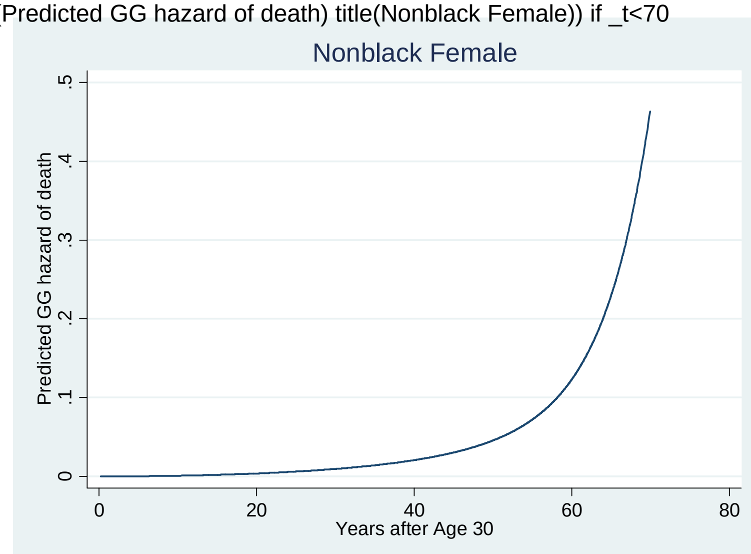
sts graph if Female==1 & Black==0, surv xtitle(Years after Age 30) ytitle(% Alive) ///
title(Nonblack Female) addplot(line GGSurvivalNonblackFemale ageexitm30, sort)



Very good fit for Non-blacks

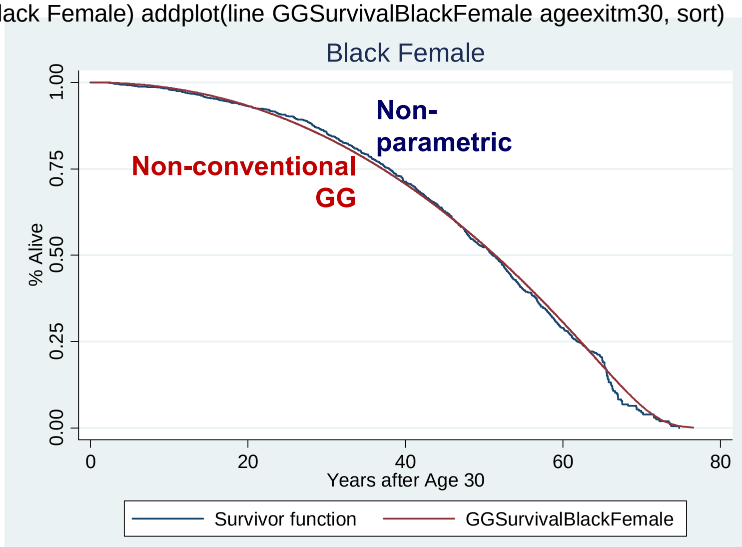
Hazard of death for non-black females

```
graph twoway (line hazardNonblackFemale _t, sort xtitle(Years after Age 30) ///
ytitle(Predicted GG hazard of death) title(Nonblack Female)) if _t<70
```



Fit of non-conventional GG to non-parametric survival for Black Females

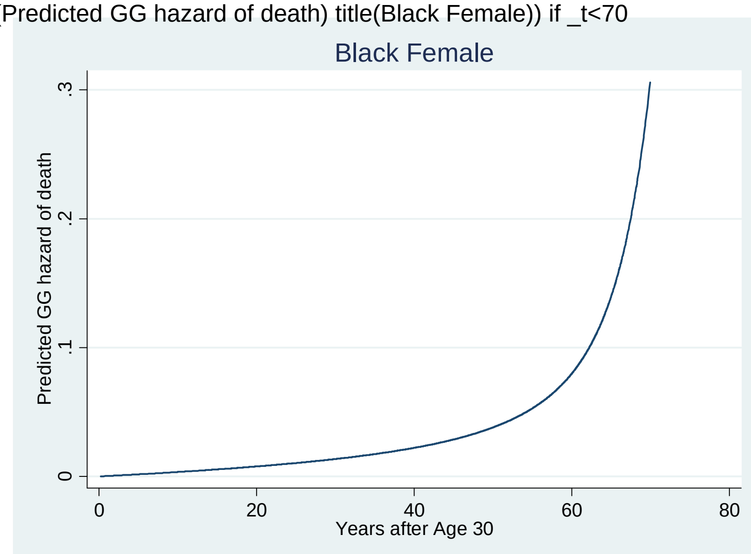
```
sts graph if Female==1 & Black==1, surv xtitle(Years after Age 30) ytitle(% Alive) ///
title(Black Female) addplot(line GGSurvivalBlackFemale ageexitm30, sort)
```



Very good fit for Blacks too

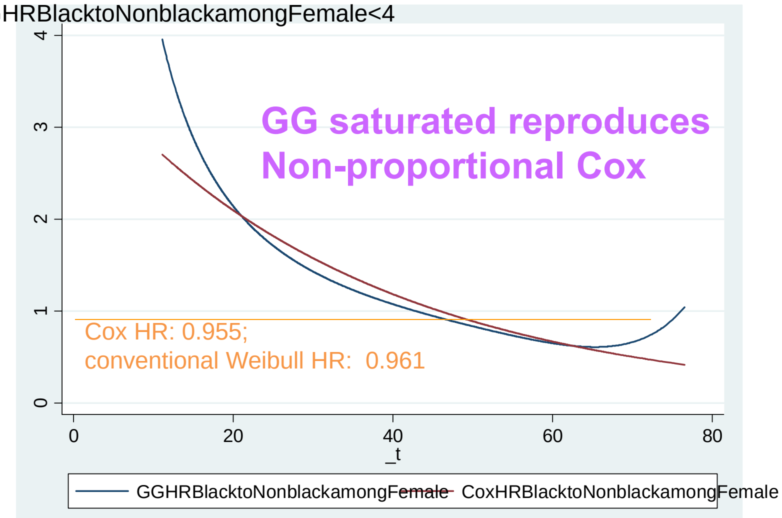
Hazard of death for black females

graph twoway (line hazardBlackFemale _t, sort xtitle(Years after Age 30) ytitle(Predicted GG hazard of death) title(Black Female)) if _t<70



Relative [black:non-black] hazard of death among female

graph twoway (line GGHRBlacktoNonblackamongFemale _t, sort) /// (line CoxHRBlacktoNonblackamongFemale _t, sort) if GGHRBlacktoNonblackamongFemale<4



Parametric Models: quantiles, not just hazards

Quantiles of GG(b,s,k) in Stata

The quantiles of any GG distribution can be easily obtained from the quantiles of a standard GG distribution with b=0, s=1, k.
From section 2.1.2 of Cox/Muñoz SIM 2007, $t_{GG(\beta,\sigma,\kappa)}(\mathrm{p})$
the pth percentile of $GG(\beta,\sigma,\kappa)$ is

$$t_{GG(\beta,\sigma,\kappa)}(\mathrm{p}) = e^{\beta}[t_{GG(0,1,\kappa)}(\mathrm{p})]^{\sigma}$$

where $t_{GG(0,1,\kappa)}(\mathrm{p})$, the pth percentile of $GG(0,1,\kappa)$ is

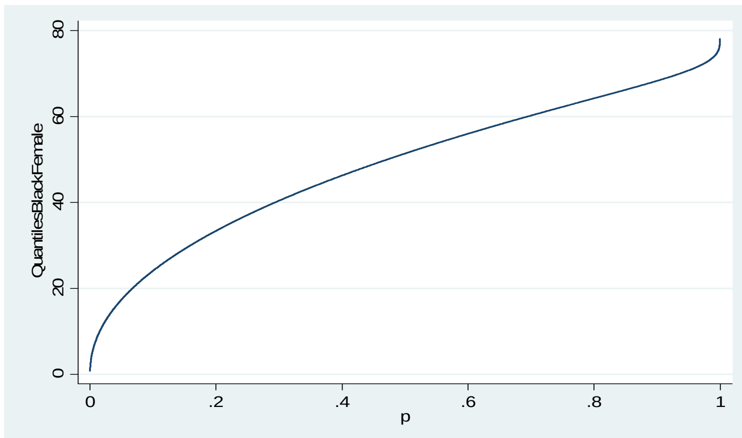
$$[\kappa^2\Gamma^{-1}(p*(\kappa>0)+(1-p)*(\kappa<0);\kappa^{-2})]^{1/\kappa}$$

*in Stata, the pth quantile of GG(b,s,k) is

exp(b)*((k^2)*invgammap(k^(-2),p*(k>0)+(1-p)*(k<0)))^(s/k)

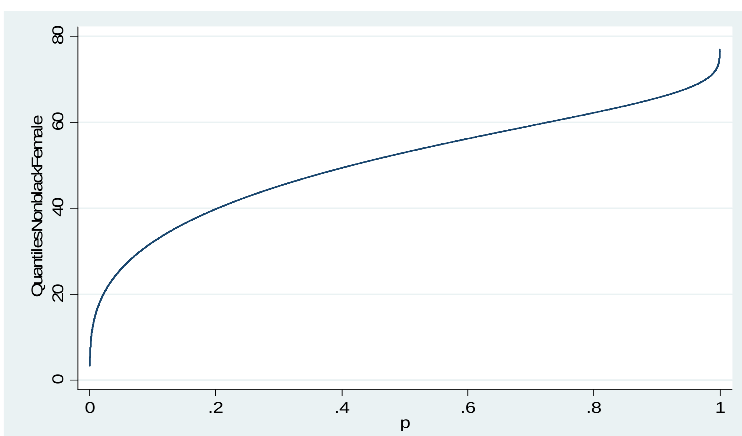
Quantiles after age 30 for Black Females

line QuantilesBlackFemale p



Quantiles after age 30 for Non-black Females

```
egen samplesize= max( _n )
gen p= _n / samplesize /*trick to create a sequence of percentages deceased*/
gen QuantilesNonblackFemale=
        exp([_t]_cons)*((e(kappa)^2)*invgammap(e(kappa)^(-2),
                (e(kappa)>0)*p+(e(kappa)<=0)*(1-p)))^(e(sigma)/e(kappa))
```



Differences# (Black - Nonblack) of Quantiles among Females

```
streg Black if Female==1, anc(Black) anc2(Black) dist(gamma)

*from section 2.1.2 of Cox/Muñoz SIM 2007
*pth quantile of GG(b,s,k) is
   exp(b)*((k^2)*invgammap(k^(-2),p*(k>0)+(1-p)*(k<0)))^(s/k)

*95% CI for difference in 20th
nlcom (twentieth: exp([_t]_cons+[_t]Black)* ///
((([kappa]_cons+[kappa]Black)^2)*invgammap(([kappa]_cons+[kap
pa]Black)^(-2),0.20)) ///
^(exp([ln_sig]_cons+[ln_sig]Black)/([kappa]_cons+[kappa]Black)) ///
- exp([_t]_cons)* ///
((([kappa]_cons)^2)*invgammap(([kappa]_cons)^(-2),0.20)) ///
^(exp([ln_sig]_cons)/([kappa]_cons)))
```

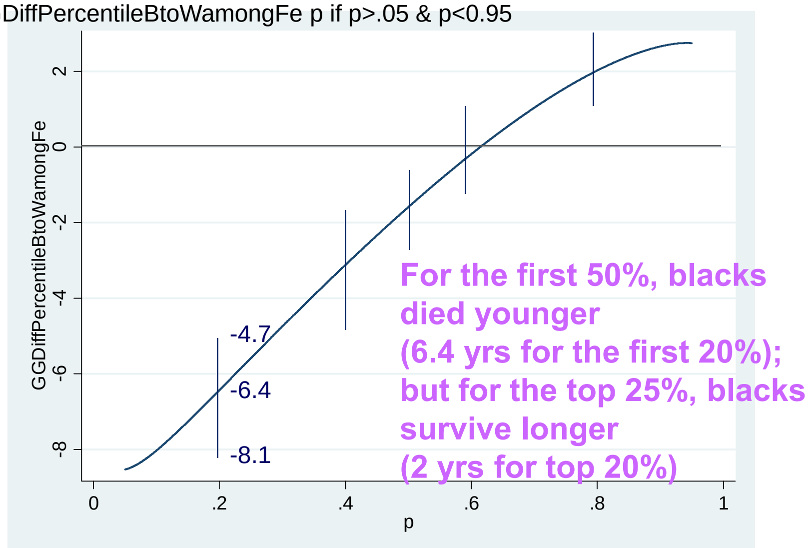
except at zero, differences can't ever be constant

Differences# (Black - Nonblack) of Quantiles among Females

```
------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   twentieth |  -6.426571   .8612467    -7.46   0.000    -8.114584   -4.738558
------------------------------------------------------------------------------
```

Differences (Black - Nonblack) of Quantiles among Females

gen GGDiffPercentileBtoWamongFe= QuantilesBlackFemale-QuantilesNonblackFemale
. line GGDiffPercentileBtoWamongFe p if p>.05 & p<0.95



Summary

Analysis of racial differences in survival among women

Thumbs down: proportional hazards and conventional (proportional times) parametric models.

Thumbs up: saturated GG which reproduces Cox's time dependent hazard ratios and more importantly, it allows for quantification of years lost/gained.

**Thumbs up: Stata's streg with anc and anc2 to incorporate non-proportionalities;
and nlcom to quantify years lost/gained.**

Differences (Female - Male) of Quantiles among Non-blacks