

The wald statistic (z column) evaluates, whether the beta (β) coefficient of a given variable is statistically significantly different from 0.

 A positive sign for the regression coefficient (coef) means that the hazard (risk of death) is higher, and thus the prognosis worse, for subjects with higher values of that variable. The variable sex is encoded as a letter vector. M: male, F: female. The R summary for the Cox model gives the hazard ratio (HR) for the second group relative to the first group, that is, male versus female. The beta coefficient for sexM = -1.16936 indicates that males have lower risk of death (lower survival rates) than females, in these data.

*Hazard ratios*. The exponentiated coefficients (exp(coef) = exp(-1.16936) = 0.31057), also known as *hazard ratios*, give the effect size of covariates. For example, being male (sex=M) reduces the hazard by a factor of 0.31057, or 69%. Being male is associated with good prognostic.

*Confidence intervals of the hazard ratios*. The summary output also gives upper and lower 95% confidence intervals for the hazard ratio (exp(coef)), lower 95% bound = 0.2992, upper 95% bound = 0.3224.

*Global statistical significance of the model*. Finally, the output gives p-values for three alternative tests for overall significance of the model: The likelihood-ratio test, Wald test, and score logrank statistics. These three methods are asymptotically equivalent. For large enough N, they will give similar results. For small N, they may differ somewhat. The Likelihood ratio test has better behavior for small sample sizes, so it is generally preferred.

Consider that, we want to assess the impact of the sex on the estimated survival probability. In this case, we construct a new data frame with two rows, one for each value of sex; the other covariates are fixed to their average values (if they are continuous variables) or to their lowest level (if they are discrete variables). For a dummy covariate, the average value is the proportion coded 1 in the data set. This data frame is passed to *survfit*() via the *newdata* argument:

# Create the new data sex\_df <- with(lung, data.frame(sex = c(1, 2), age = rep(mean(age, na.rm = TRUE), 2), ph.ecog = c(1, 1) ) ) sex\_df

sex age ph.ecog 1 1 62.44737 1 2 2 62.44737 1

# Survival curves fit <- survfit(res.cox, newdata = sex\_df) ggsurvplot(fit, conf.int = TRUE, legend.labs=c("Sex=1", "Sex=2"), ggtheme = theme\_minimal())

Using the tutorial, I began by looking into the effect of sex with univariate cox proportional hazard model. Censoring or not censoring, I found that there was significant correlation with time and it did not pass the *Schoenfeld residuals* test. Possible solutions include adding a covariate\*time interaction or stratification. I will try first adding in other factors that I believe should have a significant impact and see if that fixes my issue. Try adding radiation treatment group.

Using coxph(Surv(age, status) ~ radn, data=demographics)) I see that Gamma is 32% worse than control and neutron is 19% worse than control. It makes sense that the irradiated animals would be worse off, but it doesn’t make sense that sex has a larger impact on survival than treatment group. Without censoring, gamma is 35% worse and neutrons are 31% worse than controls. Clearly, sensoring is having a big impact. I should first remove some data that needs to be removed (JM11 and neutron data) and then begin my real analysis comparing cox proportional hazard, KM, and with or without censoring.

Starting the analysis from the beginning…

* combine demographics and macro
* delete JM11 (not real data set) and JM10 (peromyscus)
* delete mice irradiated with neutrons
* delete mice from JM14 treated with radio protectants
* check for mouse that has lethal cause of death
  + 40366 have lethal cause of death, 9744 do not have a listed cause of death

|  |  |  |
| --- | --- | --- |
| **Data removed** | **Reasoning** | **Total # of mice** |
| - | - | 50110 |
| JM11 | Not a real data set | 49225 |
| JM10 | peromysucs | 46835 |
| Neutron irradiated mice | Testing for dose rate effects, neutrons | 25425 |
| JM14 mice treated with radioprotectors | We are not interested in the effects of radioprotectors | 24225 |
| Breeder mice | Held under different conditions | 24107 |
| JM2 mice | Held under different conditions | 17317 |
| COD – Removal to another experiment | Mice listed under different experiment, do not want to double count mice | 15137 |

|  |  |
| --- | --- |
| **Mice censored** | **Number of mice** |
| COD – Accidental death | 47 |
| COD – Escaped during irradiation | 8 |
| COD – Discarded | 207 |
| COD – Improper irradiation | 77 |
| COD – Missing | 29 |
| COD – Sacrifice, programmed | 19 |
| No lethal disease listed and not already censored | 936 |
| **Total censored mice:** | **1323** |

Now eliminate data based on survival curves… compare KM and Cox, both with and without censoring to see how it affects everything.

* 1. KM - use all data with total\_dose == 0, stratify by expt, no censoring:

N Observed Expected (O-E)^2/E (O-E)^2/V

expt=3 385 385 360 1.7948 1.9475

expt=4 1071 1071 1028 1.7960 2.2586

expt=7 490 490 499 0.1612 0.1796

expt=8 110 110 108 0.0528 0.0542

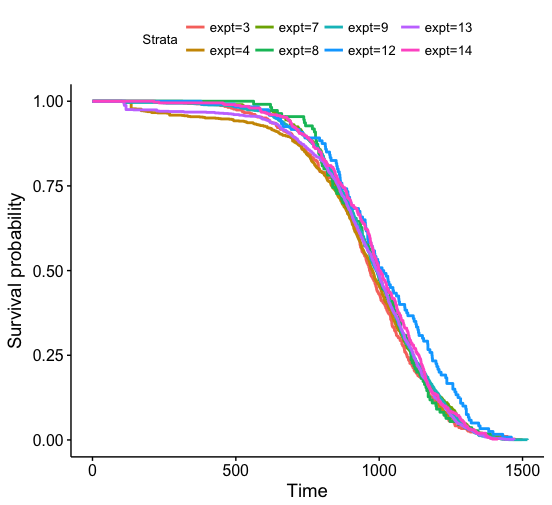
expt=9 1350 1350 1370 0.3029 0.4158

expt=12 120 120 155 7.7319 8.0481

expt=13 1214 1214 1200 0.1690 0.2218

expt=14 400 400 421 1.0556 1.1569

Chisq= 13.2 on 7 degrees of freedom, p= 0.0678



b. Use all data with total\_dose == 0 and censor

N Observed Expected (O-E)^2/E (O-E)^2/V

expt=3 385 362 328.6 3.390463 3.678338

expt=4 1071 981 940.9 1.711701 2.151858

expt=7 490 457 457.3 0.000209 0.000233

expt=8 110 98 98.4 0.001730 0.001777

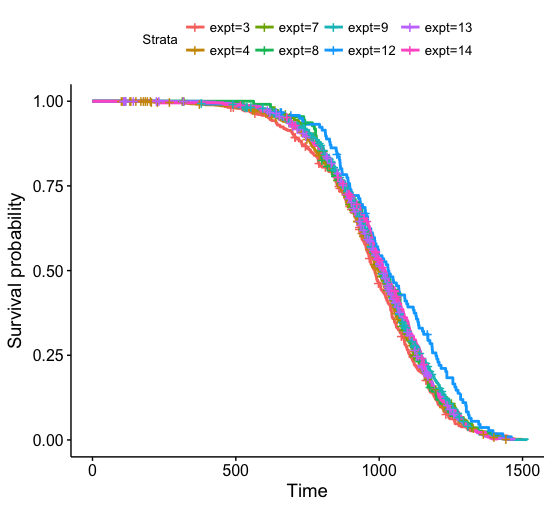
expt=9 1350 1215 1255.4 1.297756 1.781378

expt=12 120 112 142.8 6.655726 6.929390

expt=13 1214 1106 1098.5 0.050872 0.066763

expt=14 400 377 386.1 0.213001 0.233441

Chisq= 13.4 on 7 degrees of freedom, p= 0.0622



1. Cox model – no censoring

graphs <- cox.survival(stratified.by = "expt", df = subset(all.info, total\_dose == 0))

Call:

coxph(formula = f, data = df)

n= 5140, number of events= 5140

coef exp(coef) se(coef) z Pr(>|z|)

expt4 -0.02719 0.97318 0.05950 -0.457 0.64769

expt7 -0.08621 0.91740 0.06819 -1.264 0.20612

expt8 -0.04551 0.95551 0.10819 -0.421 0.67401

expt9 -0.08316 0.92020 0.05788 -1.437 0.15080

expt12 -0.32317 0.72385 0.10467 -3.088 0.00202 \*\*

expt13 -0.05619 0.94536 0.05858 -0.959 0.33754

expt14 -0.11974 0.88715 0.07152 -1.674 0.09408 .

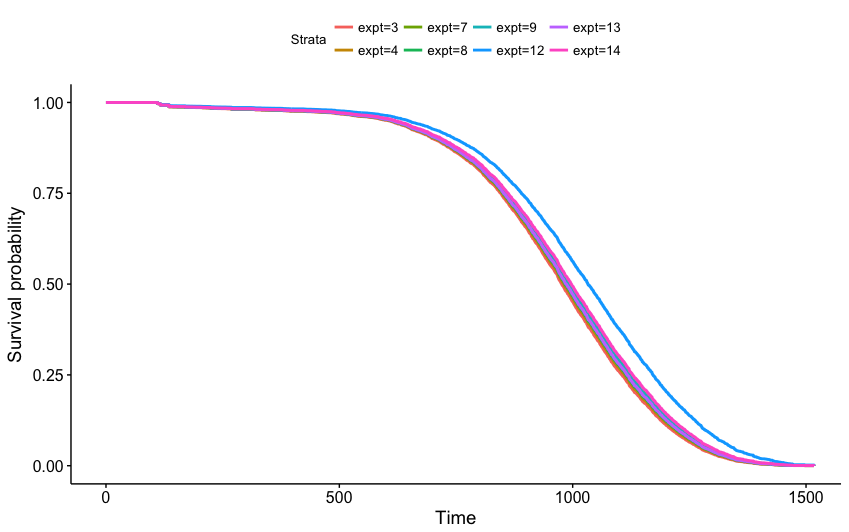
Concordance= 0.516 (se = 0.005 )

Rsquare= 0.003 (max possible= 1 )

Likelihood ratio test= 13.78 on 7 df, p=0.05532

Wald test = 13.11 on 7 df, p=0.06952

Score (logrank) test = 13.16 on 7 df, p=0.06829



rho chisq p

expt4 -0.000433 0.000966 0.9752

expt7 0.011547 0.686050 0.4075

expt8 0.017238 1.528411 0.2164

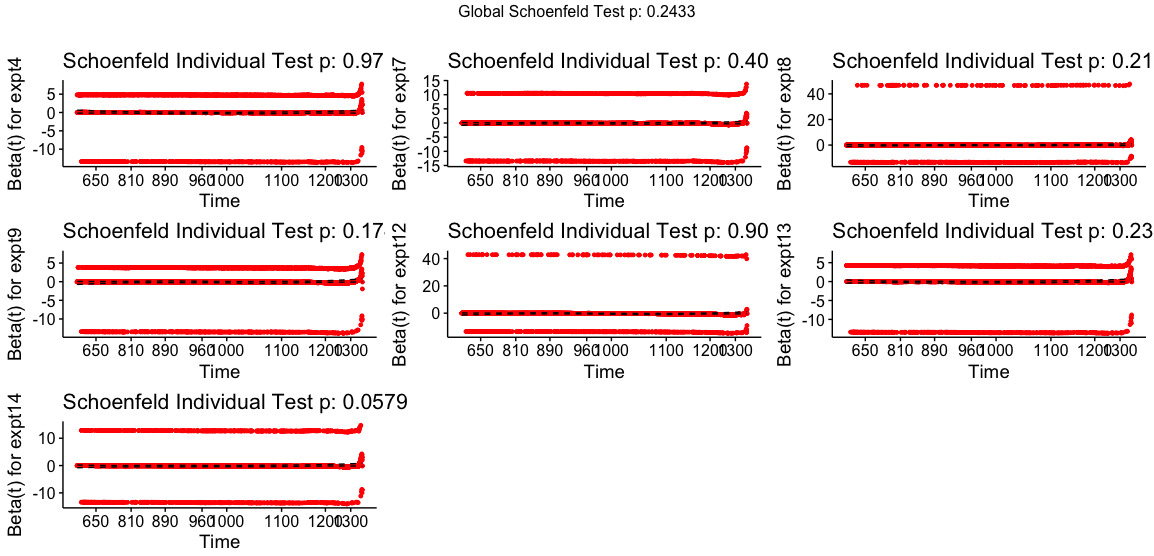
expt9 0.018905 1.840272 0.1749

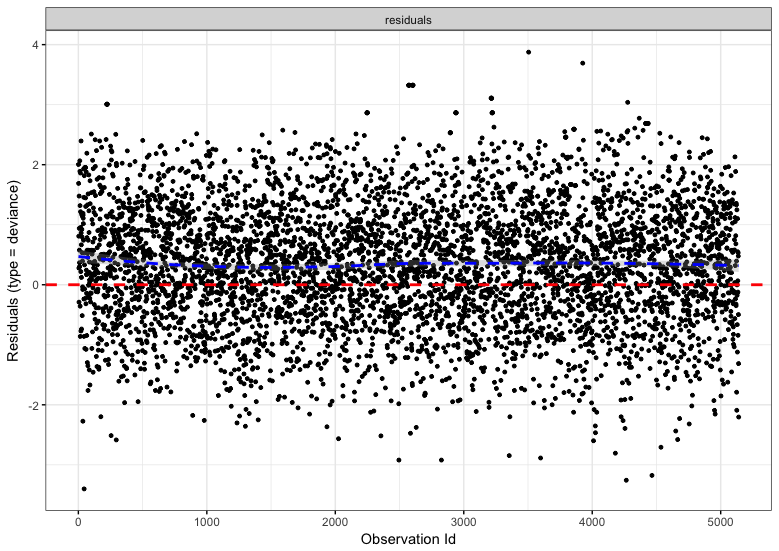
expt12 0.001631 0.013641 0.9070

expt13 0.016698 1.435380 0.2309

expt14 0.026427 3.596739 0.0579

GLOBAL NA 9.132278 0.2433





1. Cox model with censoring

cox.censored(stratified.by = "expt", df = subset(all.info, total\_dose == 0))

Call:

coxph(formula = f, data = df)

n= 5140, number of events= 4708

coef exp(coef) se(coef) z Pr(>|z|)

expt4 -0.05495 0.94653 0.06158 -0.892 0.3722

expt7 -0.09736 0.90723 0.07045 -1.382 0.1670

expt8 -0.10035 0.90452 0.11395 -0.881 0.3785

expt9 -0.12957 0.87847 0.06000 -2.160 0.0308 \*

expt12 -0.34169 0.71057 0.10825 -3.157 0.0016 \*\*

expt13 -0.08983 0.91408 0.06065 -1.481 0.1386

expt14 -0.12067 0.88633 0.07372 -1.637 0.1017

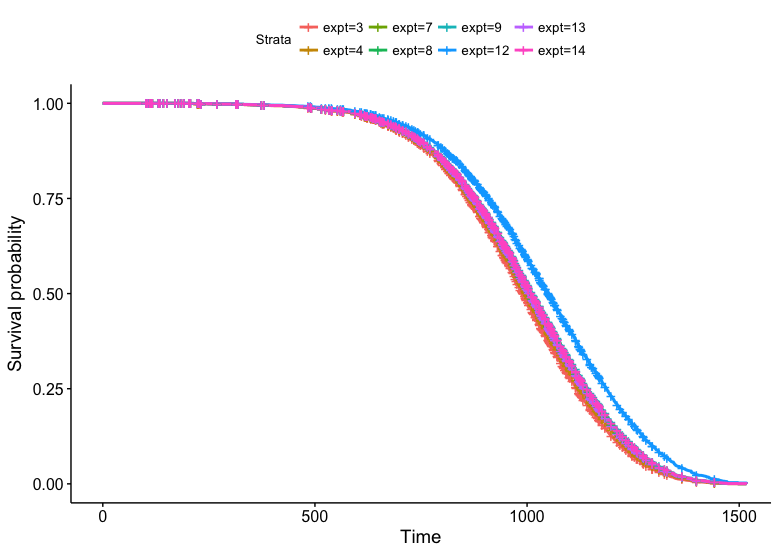
Concordance= 0.515 (se = 0.005 )

Rsquare= 0.003 (max possible= 1 )

Likelihood ratio test= 13.83 on 7 df, p=0.05422

Wald test = 13.36 on 7 df, p=0.06387

Score (logrank) test = 13.4 on 7 df, p=0.06286



rho chisq p

expt4 0.01260 0.748 0.3873

expt7 0.01049 0.518 0.4715

expt8 0.01768 1.472 0.2250

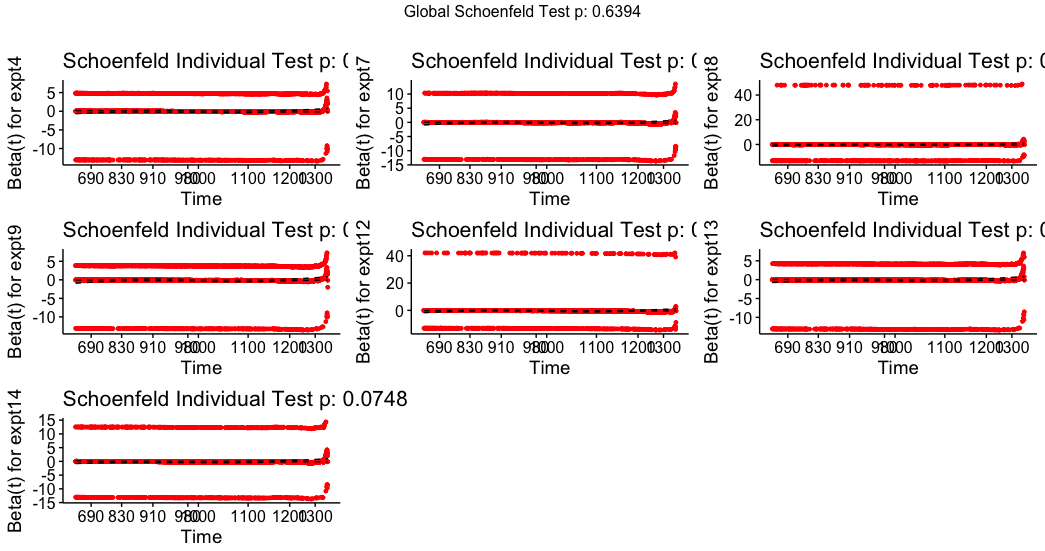
expt9 0.02015 1.916 0.1663

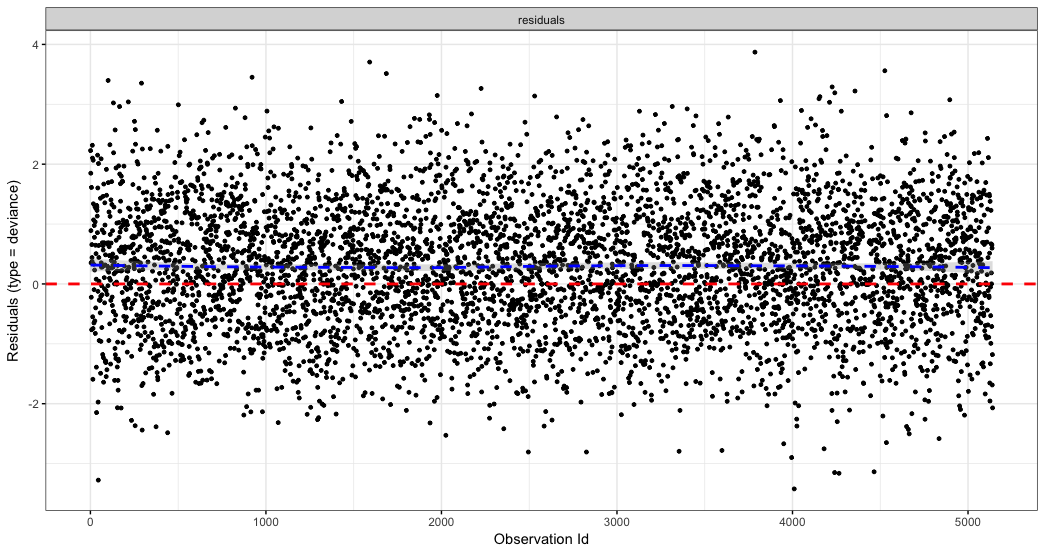
expt12 0.00889 0.371 0.5424

expt13 0.02424 2.771 0.0960

expt14 0.02594 3.174 0.0748

GLOBAL NA 5.169 0.6394





Not much to say about the differences between models yet, but I can say that Cox PH model works (no signifance in residuals test and outliers test looks symmetric around 0). We can also say that JM12 is significantly different from both Cox PH models. Remove it and repeat.

1. Remove JM 12 and repeat
   1. km.survival(stratified.by = "expt", df = subset(all.info, total\_dose == 0))

N Observed Expected (O-E)^2/E (O-E)^2/V

expt=3 385 385 362 1.4546 1.5829

expt=4 1071 1071 1035 1.2496 1.5838

expt=7 490 490 502 0.3100 0.3467

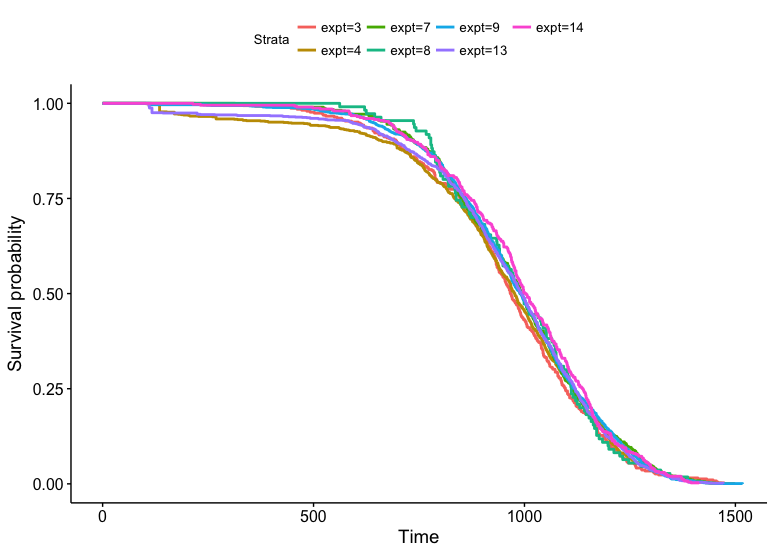
expt=8 110 110 108 0.0256 0.0263

expt=9 1350 1350 1380 0.6546 0.9089

expt=13 1214 1214 1208 0.0303 0.0402

expt=14 400 400 424 1.3689 1.5046

Chisq= 5.1 on 6 degrees of freedom, p= 0.526



* 1. km.censored(stratified.by = "expt", df = subset(all.info, total\_dose == 0))

N Observed Expected (O-E)^2/E (O-E)^2/V

expt=3 385 362 330.8 2.94e+00 3.20e+00

expt=4 1071 981 947.1 1.21e+00 1.54e+00

expt=7 490 457 460.4 2.58e-02 2.88e-02

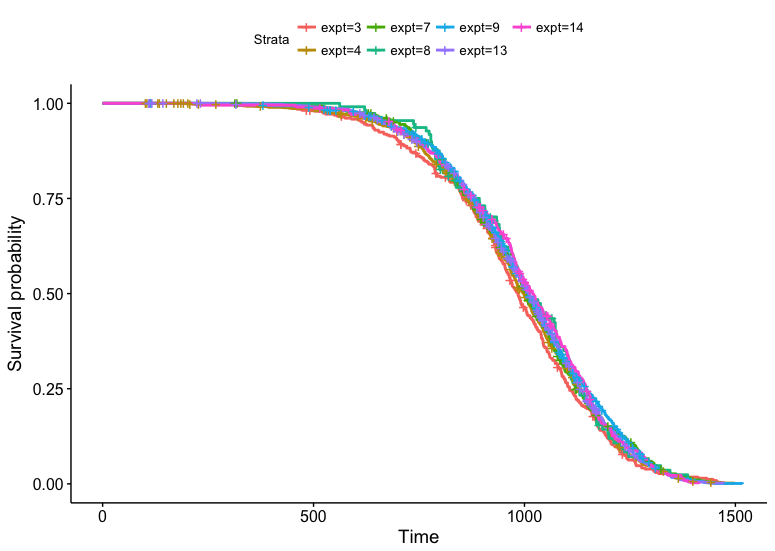
expt=8 110 98 99.1 1.12e-02 1.15e-02

expt=9 1350 1215 1264.0 1.90e+00 2.64e+00

expt=13 1214 1106 1105.8 3.03e-05 4.01e-05

expt=14 400 377 388.8 3.55e-01 3.91e-01

Chisq= 6.5 on 6 degrees of freedom, p= 0.369



* 1. graphs <- cox.survival(stratified.by = "expt", df = subset(all.info, total\_dose == 0))

n= 5020, number of events= 5020

coef exp(coef) se(coef) z Pr(>|z|)

expt4 -0.02726 0.97311 0.05951 -0.458 0.6469

expt7 -0.08643 0.91720 0.06819 -1.267 0.2050

expt8 -0.04553 0.95549 0.10819 -0.421 0.6739

expt9 -0.08349 0.91990 0.05789 -1.442 0.1492

expt13 -0.05632 0.94524 0.05859 -0.961 0.3364

expt14 -0.12018 0.88676 0.07153 -1.680 0.0929 .

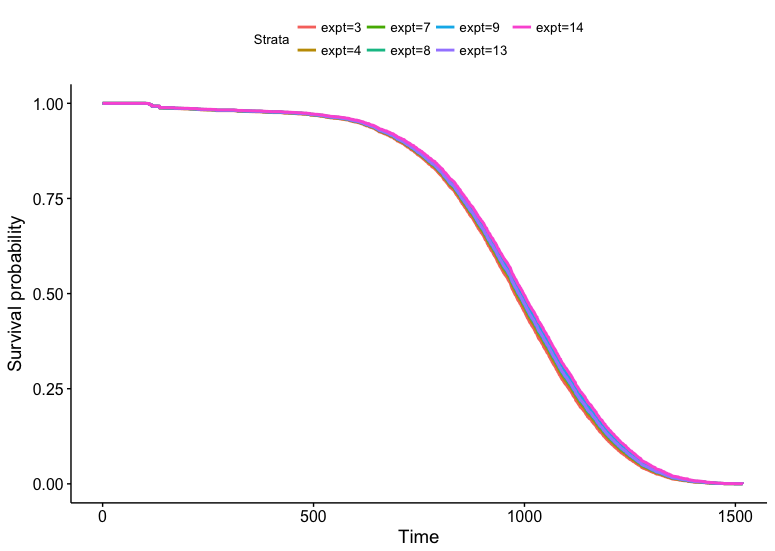
Concordance= 0.514 (se = 0.005 )

Rsquare= 0.001 (max possible= 1 )

Likelihood ratio test= 5.08 on 6 df, p=0.5331

Wald test = 5.09 on 6 df, p=0.5322

Score (logrank) test = 5.09 on 6 df, p=0.5319



rho chisq p

expt4 -0.000478 0.00115 0.9730

expt7 0.011643 0.68127 0.4091

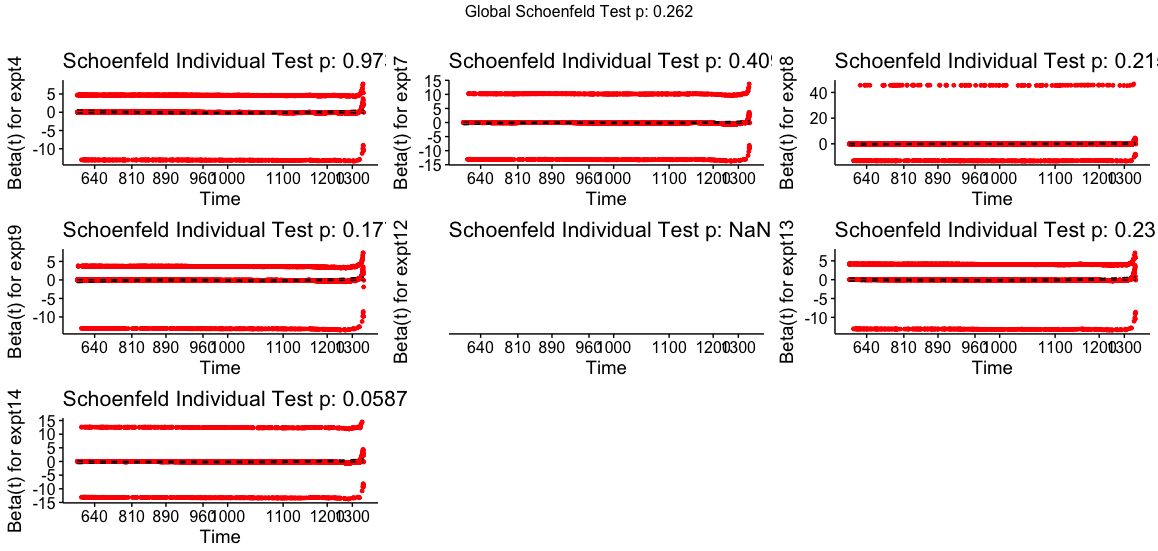
expt8 0.017491 1.53701 0.2151

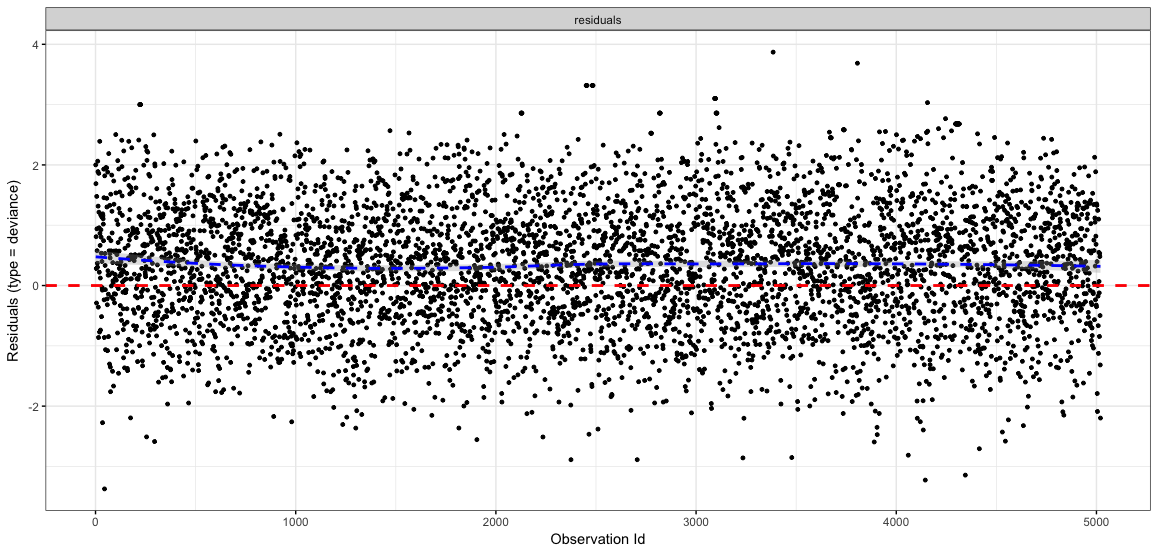
expt9 0.019025 1.82034 0.1773

expt13 0.016863 1.42985 0.2318

expt14 0.026652 3.57322 0.0587

GLOBAL NA 8.87105 0.2620





* 1. cox.censored(stratified.by = "expt", df = subset(all.info, total\_dose == 0))

n= 5020, number of events= 4596

coef exp(coef) se(coef) z Pr(>|z|)

expt4 -0.05503 0.94646 0.06159 -0.894 0.3716

expt7 -0.09759 0.90702 0.07046 -1.385 0.1660

expt8 -0.10040 0.90447 0.11395 -0.881 0.3783

expt9 -0.12993 0.87816 0.06000 -2.165 0.0304 \*

expt13 -0.08998 0.91395 0.06066 -1.483 0.1380

expt14 -0.12112 0.88593 0.07373 -1.643 0.1004

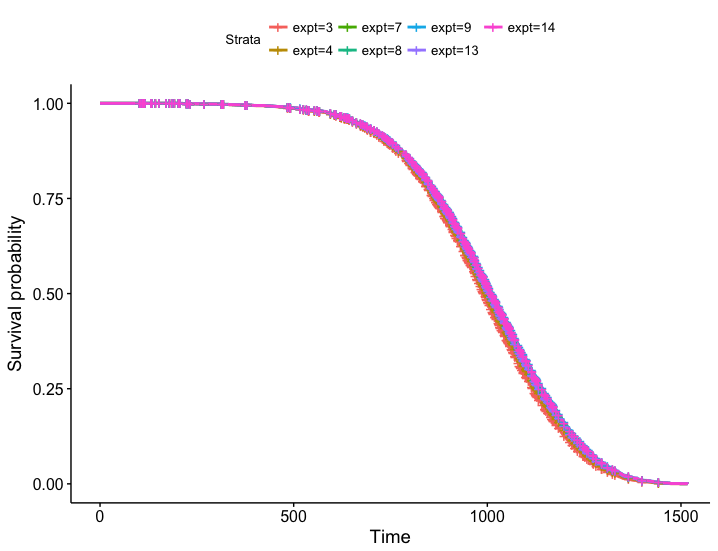
Concordance= 0.513 (se = 0.005 )

Rsquare= 0.001 (max possible= 1 )

Likelihood ratio test= 6.38 on 6 df, p=0.3822

Wald test = 6.45 on 6 df, p=0.3751

Score (logrank) test = 6.45 on 6 df, p=0.3746



rho chisq p

expt4 0.0127 0.746 0.3877

expt7 0.0106 0.514 0.4732

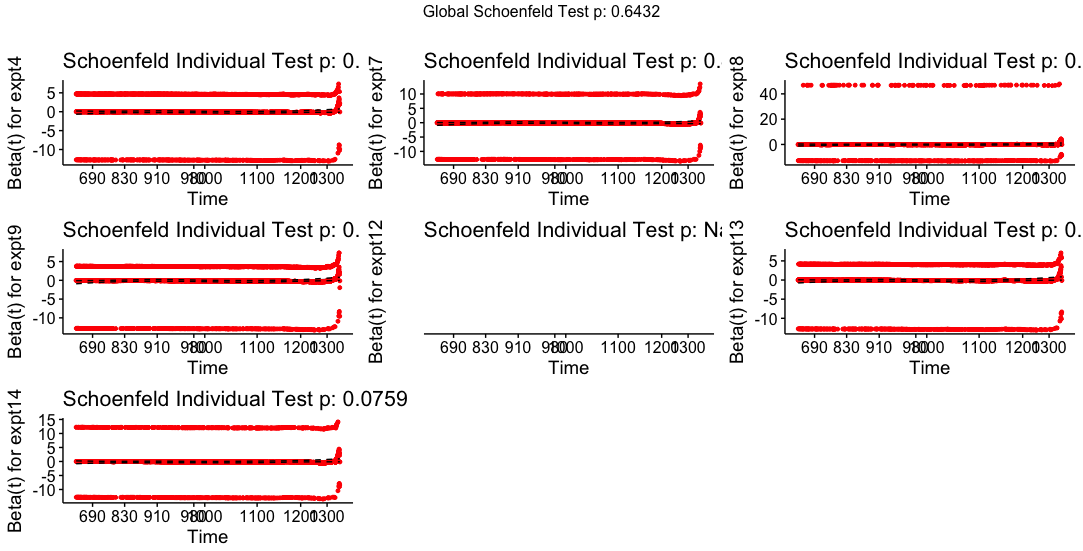
expt8 0.0179 1.480 0.2238

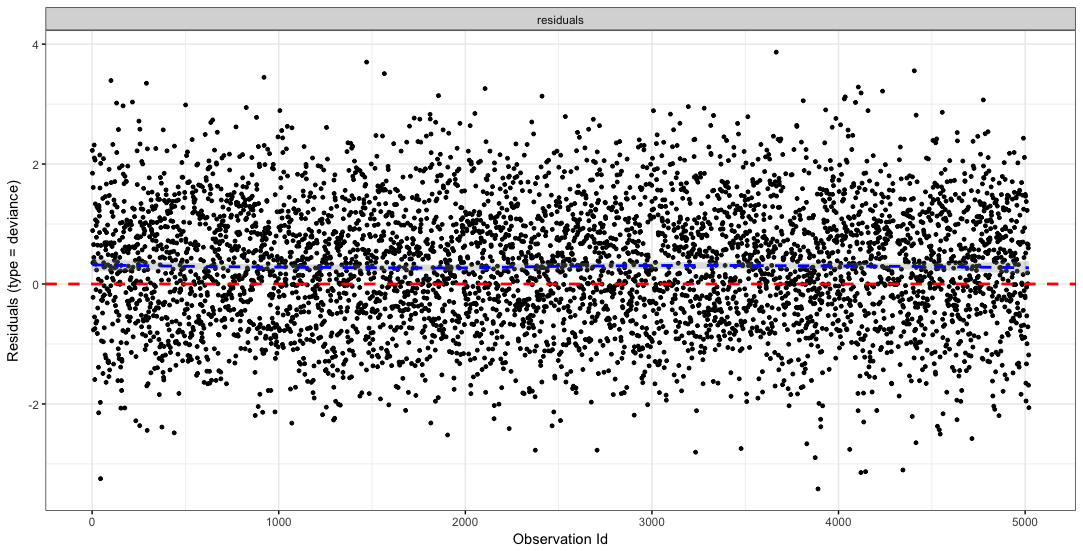
expt9 0.0203 1.893 0.1688

expt13 0.0245 2.767 0.0962

expt14 0.0262 3.151 0.0759

GLOBAL NA 5.137 0.6432





Conclusions:

The overall model is not significant for any tests (likelihood ratio, wald, nor score) meaning b1=b2=...=bn=0. That means the model is the baseline hazard function, only a function of time and not a function of x. This should be what we want for controls! Not dependent on the experiment, which is our variable.

Expt 9 becomes significantly different from the reference (expt 3) when censoring… investigate why. How many are censored? What kind of data are we dealing with? What data did I include when I only used KM curves?

Originally excluded JM9 males when I did this analysis. Maybe they can be included with new analysis method? JM 9 summary

sex radn status fractions n mean\_age

<fctr> <fctr> <dbl> <int> <int> <dbl>

1 F C 0 1 110 880.0545

2 F C 0 24 14 1002.2857

3 F C 1 1 840 984.1226

4 F C 1 24 186 968.3871

5 F G 0 1 106 855.5189

6 F G 1 1 944 964.5318

7 M C 0 1 11 979.7273

8 M C 1 1 189 1046.2222

Should have been including sex as a covariate… sex has significant impact on death and there may be more or less males/females in each experiment. Redo… maybe can keep JM12. Only use Cox PH model now though.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* REDO with sex as covariate \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

graphs <- cox.censored(stratified.by = c("sex","expt"), df = subset(all.info, total\_dose == 0))

n= 5140, number of events= 4708

coef exp(coef) se(coef) z Pr(>|z|)

sexM -0.10237 0.90269 0.03146 -3.254 0.00114 \*\*

expt4 -0.05415 0.94729 0.06158 -0.879 0.37916

expt7 -0.09165 0.91242 0.07047 -1.301 0.19339

expt8 -0.09236 0.91177 0.11397 -0.810 0.41771

expt9 -0.16758 0.84571 0.06109 -2.743 0.00608 \*\*

expt12 -0.29698 0.74306 0.10917 -2.720 0.00652 \*\*

expt13 -0.09562 0.90881 0.06067 -1.576 0.11503

expt14 -0.12495 0.88254 0.07373 -1.695 0.09012 .

Concordance= 0.52 (se = 0.005 )

Rsquare= 0.005 (max possible= 1 )

Likelihood ratio test= 24.44 on 8 df, p=0.001931

Wald test = 24.06 on 8 df, p=0.002237

Score (logrank) test = 24.12 on 8 df, p=0.002187

rho chisq p

sexM 0.00308 0.044 0.8339

expt4 0.01246 0.731 0.3927

expt7 0.00939 0.415 0.5194

expt8 0.01857 1.626 0.2022

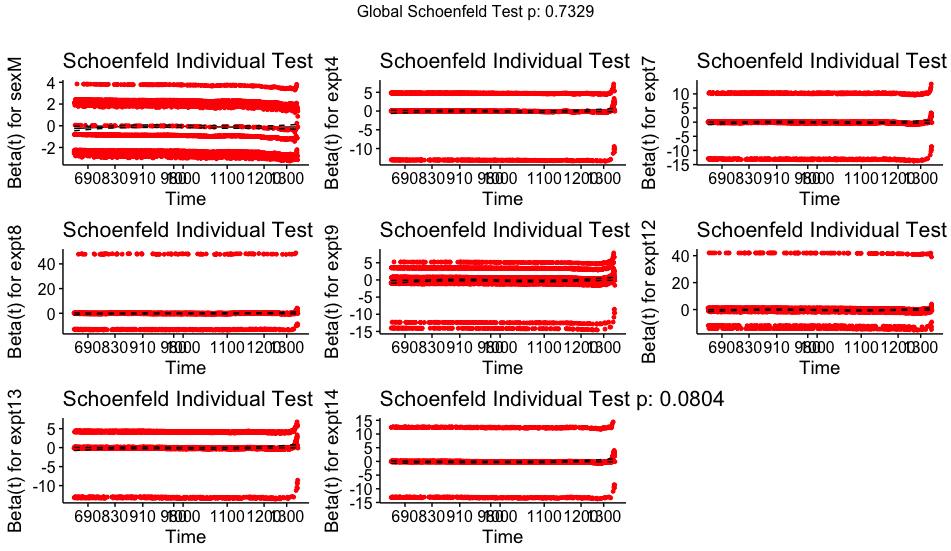
expt9 0.02051 1.980 0.1594

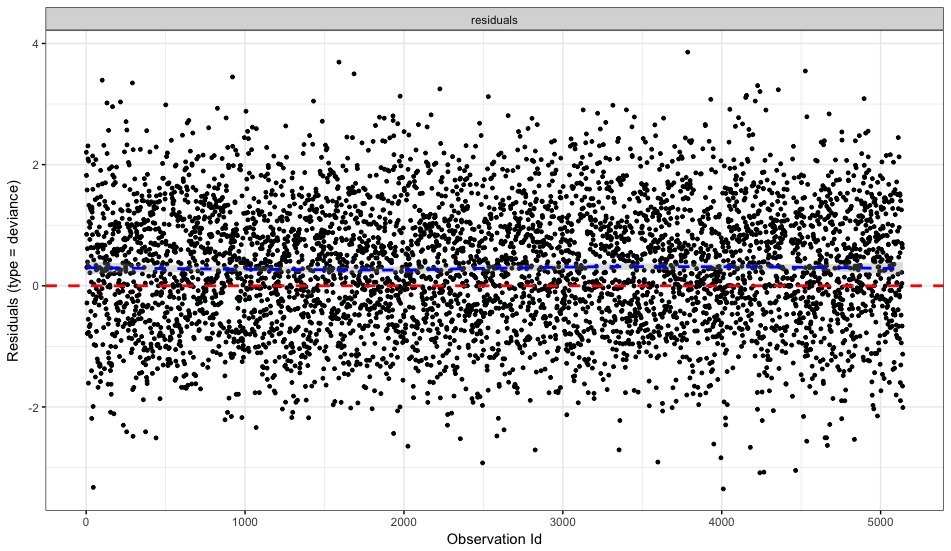
expt12 0.00793 0.295 0.5872

expt13 0.02368 2.646 0.1038

expt14 0.02546 3.057 0.0804

GLOBAL NA 5.228 0.7329





Sex is significant and makes the whole model significan. Expt 12 is still an issue, expt 9 is too, maybe remove 12 and see if there’s a differene. Then remove expt 9 males like last time I analyzed the data.

Expt 12 still an issue – remove it!

graphs <- cox.censored(stratified.by = c("sex","expt"), df = subset(all.info, total\_dose == 0))

Surv(age, status) ~ sex + expt

<environment: 0x1107dd840>

Call:

coxph(formula = f, data = df)

n= 5020, number of events= 4596

coef exp(coef) se(coef) z Pr(>|z|)

sexM -0.10238 0.90269 0.03146 -3.254 0.00114 \*\*

expt4 -0.05422 0.94722 0.06158 -0.880 0.37859

expt7 -0.09188 0.91221 0.07047 -1.304 0.19232

expt8 -0.09239 0.91175 0.11398 -0.811 0.41760

expt9 -0.16795 0.84540 0.06109 -2.749 0.00598 \*\*

expt13 -0.09576 0.90868 0.06068 -1.578 0.11454

expt14 -0.12540 0.88215 0.07374 -1.701 0.08901 .

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

exp(coef) exp(-coef) lower .95 upper .95

sexM 0.9027 1.108 0.8487 0.9601

expt4 0.9472 1.056 0.8395 1.0687

expt7 0.9122 1.096 0.7945 1.0473

expt8 0.9117 1.097 0.7292 1.1400

expt9 0.8454 1.183 0.7500 0.9529

expt13 0.9087 1.100 0.8068 1.0234

expt14 0.8821 1.134 0.7634 1.0193

Concordance= 0.518 (se = 0.005 )

Rsquare= 0.003 (max possible= 1 )

Likelihood ratio test= 16.99 on 7 df, p=0.01748

Wald test = 17.09 on 7 df, p=0.01683

Score (logrank) test = 17.1 on 7 df, p=0.01679

rho chisq p

sexM 0.00315 0.0448 0.8324

expt4 0.01260 0.7297 0.3930

expt7 0.00946 0.4114 0.5213

expt8 0.01885 1.6354 0.2010

expt9 0.02065 1.9589 0.1616

expt13 0.02396 2.6440 0.1039

expt14 0.02567 3.0347 0.0815

GLOBAL NA 5.1924 0.6365

Expt 9 still an issue. Remove the males.

graphs <- cox.censored(stratified.by = c("sex","expt"), df = subset(all.info, total\_dose == 0 & !(expt == 9 & sex == "M")))

Surv(age, status) ~ sex + expt

<environment: 0x1118bb150>

Call:

coxph(formula = f, data = df)

n= 4820, number of events= 4407

coef exp(coef) se(coef) z Pr(>|z|)

sexM -0.06992 0.93247 0.03466 -2.017 0.0437 \*

expt4 -0.05593 0.94560 0.06158 -0.908 0.3637

expt7 -0.09523 0.90917 0.07048 -1.351 0.1767

expt8 -0.09672 0.90781 0.11398 -0.849 0.3961

expt9 -0.12519 0.88233 0.06415 -1.951 0.0510 .

expt13 -0.09545 0.90897 0.06068 -1.573 0.1157

expt14 -0.12610 0.88152 0.07373 -1.710 0.0872 .

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

exp(coef) exp(-coef) lower .95 upper .95

sexM 0.9325 1.072 0.8712 0.998

expt4 0.9456 1.058 0.8381 1.067

expt7 0.9092 1.100 0.7919 1.044

expt8 0.9078 1.102 0.7261 1.135

expt9 0.8823 1.133 0.7781 1.001

expt13 0.9090 1.100 0.8070 1.024

expt14 0.8815 1.134 0.7629 1.019

Concordance= 0.512 (se = 0.005 )

Rsquare= 0.002 (max possible= 1 )

Likelihood ratio test= 7.96 on 7 df, p=0.3365

Wald test = 8.02 on 7 df, p=0.3307

Score (logrank) test = 8.02 on 7 df, p=0.3305

rho chisq p

sexM 0.000618 0.00168 0.9673

expt4 0.012210 0.65701 0.4176

expt7 0.009369 0.38657 0.5341

expt8 0.018493 1.50913 0.2193

expt9 0.017913 1.42725 0.2322

expt13 0.023788 2.49989 0.1139

expt14 0.025346 2.83677 0.0921

GLOBAL NA 4.82341 0.6815

Sex became much less significant, but still significant (must’ve been a big difference for expt 9) and expt 9 became insignificant (barely). Expt 14 is also barely insignificant. Also, the overall model is no longer significant (sex being less significant probably explains why).

Check what happens when I remove sex as covariate out of curiosity… prediction is that the values will change slightly because they’re not being separated out by sex when running calculations. Expt 9 became less significant, which I did not expect. Expt 9 only had female mice… must’ve changed estimates for other points and that impacted expt 9’s coefficient.

Current model that’s ok – Expt 4, 7, 8, 9 (females only), 13, 14.

Now check when stratifying by fractions as well. Start with univariate model. Remove expt 8 first!

graphs <- cox.censored(stratified.by = c("fractions"), df = subset(all.info, total\_dose == 0 & !(expt == 9 & sex == "M") & expt != 8))

Surv(age, status) ~ fractions

<environment: 0x1122f4d40>

Call:

coxph(formula = f, data = df)

n= 4710, number of events= 4309

coef exp(coef) se(coef) z Pr(>|z|)

fractions1 0.012939 1.013023 0.053148 0.243 0.808

fractions24 0.022338 1.022589 0.058475 0.382 0.702

fractions60 0.007909 1.007940 0.055635 0.142 0.887

fractions120 -0.050369 0.950879 0.087934 -0.573 0.567

fractions300 0.352135 1.422101 0.089638 3.928 8.55e-05 \*\*\*

Concordance= 0.507 (se = 0.005 )

Rsquare= 0.004 (max possible= 1 )

Likelihood ratio test= 18.31 on 5 df, p=0.002583

Wald test = 20.17 on 5 df, p=0.001163

Score (logrank) test = 20.36 on 5 df, p=0.00107

rho chisq p

fractions1 0.00755 0.246 0.620

fractions24 0.00982 0.416 0.519

fractions60 0.01392 0.835 0.361

fractions120 -0.01732 1.293 0.255

fractions300 0.00727 0.228 0.633

300 fractions gives much worse outcome for survival was 30% worse compared to 0 fractions, so these data need to be excluded.

Excluding 300 fractions:

graphs <- cox.censored(stratified.by = c("fractions"), df = subset(all.info, total\_dose == 0 & !(expt == 9 & sex == "M") & expt != 8 & fractions != 300))

Surv(age, status) ~ fractions

<environment: 0x11c64d8d0>

Call:

coxph(formula = f, data = df)

n= 4535, number of events= 4137

coef exp(coef) se(coef) z Pr(>|z|)

fractions1 0.012665 1.012745 0.053148 0.238 0.812

fractions24 0.022161 1.022408 0.058475 0.379 0.705

fractions60 0.007587 1.007616 0.055635 0.136 0.892

fractions120 -0.050057 0.951175 0.087934 -0.569 0.569

Concordance= 0.501 (se = 0.005 )

Rsquare= 0 (max possible= 1 )

Likelihood ratio test= 0.84 on 4 df, p=0.9327

Wald test = 0.83 on 4 df, p=0.9343

Score (logrank) test = 0.83 on 4 df, p=0.9343

rho chisq p

fractions1 0.00763 0.241 0.623

fractions24 0.00996 0.411 0.522

fractions60 0.01414 0.828 0.363

fractions120 -0.01768 1.293 0.255

graphs <- cox.censored(stratified.by = c("sex", "expt", "fractions"), df = subset(all.info, total\_dose == 0 & !(expt == 9 & sex == "M") & expt != 8 & fractions != 300))

Surv(age, status) ~ sex + expt + fractions

<environment: 0x111bced40>

Call:

coxph(formula = f, data = df)

n= 4535, number of events= 4137

coef exp(coef) se(coef) z Pr(>|z|)

sexM -0.09366 0.91060 0.03750 -2.498 0.0125 \*

expt4 -0.23037 0.79424 0.10486 -2.197 0.0280 \*

expt7 -0.21641 0.80541 0.13847 -1.563 0.1181

expt9 -0.15716 0.85456 0.06622 -2.373 0.0176 \*

expt13 -0.21941 0.80299 0.13453 -1.631 0.1029

expt14 -0.12755 0.88025 0.07374 -1.730 0.0837 .

fractions1 -0.12254 0.88467 0.11959 -1.025 0.3055

fractions24 -0.01133 0.98873 0.08795 -0.129 0.8975

fractions60 NA NA 0.00000 NA NA

fractions120 NA NA 0.00000 NA NA

rho chisq p

sexM 0.0159 1.02 0.3121

expt4 0.0168 1.16 0.2808

expt7 0.0255 2.68 0.1015

expt9 0.0233 2.26 0.1330

expt13 0.0328 4.45 0.0350

expt14 0.0264 2.88 0.0897

fractions1 0.0243 2.42 0.1195

fractions24 0.0310 3.93 0.0475

fractions60 NA NaN NaN

fractions120 NA NaN NaN

Not what I was expecting… why did everything get messed up?

Expt 9 and expt4 are not significant… 60 and 120 fractions mysteries.

########### PAUSE – super confused ###########

Why are my coefficients changing so much?

https://stats.stackexchange.com/questions/52067/does-adding-more-variables-into-a-multivariable-regression-change-coefficients-o

A parameter estimate in a regression model (e.g., β̂ iβ^i) will change if a variable, XjXj, is added to the model that is:

1. correlated with that parameter's corresponding variable, XiXi (which was already in the model), **and**
2. correlated with the response variable, YY

How to check for these correlations? <https://stats.stackexchange.com/questions/108007/correlations-with-categorical-variables>

**Significance tests:**

* Continuous vs. Nominal: run an [ANOVA](http://en.wikipedia.org/wiki/Anova). In R, you can use [?aov](http://stat.ethz.ch/R-manual/R-patched/library/stats/html/aov.html).
* Nominal vs. Nominal: run a [chi-squared test](http://en.wikipedia.org/wiki/Pearson%27s_chi-squared_test#Test_of_independence). In R, you use [?chisq.test](http://stat.ethz.ch/R-manual/R-patched/library/stats/html/chisq.test.html).

Chi-squared test… understand it better. https://onlinecourses.science.psu.edu/stat500/node/56

H0: In the population, the two categorical variables are independent.Ha: In the population, two categorical variables are dependent.

 Instead of using the words "independent" and "dependent" one could say "there is no relationship between the two categorical variables" versus "there is a relationship between the two categorical variables".

Example for calculating values: <http://www.stat.yale.edu/Courses/1997-98/101/chisq.htm>

My data – want to see relationships between fractions, expt, sex. Want to also check if they’re correlated to age (dependent variable).

> chisq.test(all.info$sex, all.info$expt)

Pearson's Chi-squared test

data: all.info$sex and all.info$expt

X-squared = 2142.6, df = 5, p-value < 2.2e-16

> chisq.test(all.info$sex, all.info$fractions)

Pearson's Chi-squared test

data: all.info$sex and all.info$fractions

X-squared = 1542.6, df = 5, p-value < 2.2e-16

> chisq.test(all.info$expt, all.info$fractions)

Pearson's Chi-squared test

data: all.info$expt and all.info$fractions

X-squared = 29469, df = 25, p-value < 2.2e-16

All are independent variables are highly dependent on each other. Now check with age.

> aov(age ~ expt, data = subset(all.info, total\_dose == 0))

Call:

aov(formula = age ~ expt, data = subset(all.info, total\_dose ==

0))

Terms:

expt Residuals

Sum of Squares 1225467 239104260

Deg. of Freedom 5 4904

Residual standard error: 220.8098

Estimated effects may be unbalanced

> summary(aov(age ~ expt, data = subset(all.info, total\_dose == 0)))

Df Sum Sq Mean Sq F value Pr(>F)

expt 5 1225467 245093 5.027 0.000135 \*\*\*

Residuals 4904 239104260 48757

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

> summary(aov(age ~ sex, data = subset(all.info, total\_dose == 0)))

Df Sum Sq Mean Sq F value Pr(>F)

sex 1 582481 582481 11.92 0.000559 \*\*\*

Residuals 4908 239747245 48848

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

> summary(aov(age ~ fractions, data = subset(all.info, total\_dose == 0)))

Df Sum Sq Mean Sq F value Pr(>F)

fractions 5 913354 182671 3.742 0.00221 \*\*

Residuals 4904 239416373 48821

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

> summary(aov(age ~ fractions + sex, data = subset(all.info, total\_dose == 0)))

Df Sum Sq Mean Sq F value Pr(>F)

fractions 5 913354 182671 3.763 0.00211 \*\*

sex 1 1414908 1414908 29.148 7.02e-08 \*\*\*

Residuals 4903 238001465 48542

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

> summary(aov(age ~ fractions + sex + expt, data = subset(all.info, total\_dose == 0)))

Df Sum Sq Mean Sq F value Pr(>F)

fractions 5 913354 182671 3.988 0.0013 \*\*

sex 1 1414908 1414908 30.887 2.88e-08 \*\*\*

expt 4 13581112 3395278 74.117 < 2e-16 \*\*\*

Residuals 4899 224420353 45809

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Conclusion – variables correlate with each other and with age, so each variable will have an impact on the others when it is added or removed from my analysis.

Figure out how to make sure expt and fractions are NOT significant in my Cox model.

Check that my subsetting from above is working as expected…

| **expt** | | **sex** | **radn** | | **fractions** | | **n** | | **mean\_age** | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  | | |  | |  |
| 3 | F | C | 1 | | 185 | | | 931.9405 | |
| 3 | M | C | 1 | | 200 | | | 985.3350 | |
| 4 | F | C | 24 | | 476 | | | 956.2185 | |
| 4 | F | G | 0 | | 20 | | | 135.0000 | |
| 4 | M | C | 24 | | 200 | | | 1010.2050 | |
| 4 | M | C | 120 | | 200 | | | 962.4000 | |
| 7 | F | C | 0 | | 180 | | | 980.9889 | |
| 7 | M | C | 0 | | 310 | | | 986.5097 | |
| 9 | F | C | 1 | | 950 | | | 972.0726 | |
| 9 | F | C | 24 | | 200 | | | 970.7600 | |
| 13 | F | C | 60 | | 600 | | | 963.2617 | |
| 13 | M | C | 60 | | 614 | | | 963.5049 | |
| 14 | F | C | 1 | | 200 | | | 977.7000 | |
| 14 | M | C | 1 | | 200 | | | 1010.3500 | |

Highlighted = strange… maybe censoring explains them?

| **expt** | | **sex** | | **radn** | **fractions** | | **status** | | **n** | | | **mean\_age** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  | |  |  | | |
| 3 | F | C | 1 | 0 | 14 | | 798.6429 | | |
| 3 | F | C | 1 | 1 | 171 | | 942.8538 | | |
| 3 | M | C | 1 | 0 | 9 | | 948.7778 | | |
| 3 | M | C | 1 | 1 | 191 | | 987.0576 | | |
| 4 | F | C | 24 | 0 | 33 | | 701.5152 | | |
| 4 | F | C | 24 | 1 | 443 | | 975.1919 | | |
| 4 | F | G | 0 | 0 | 20 | | 135.0000 | | |
| 4 | M | C | 24 | 0 | 15 | | 692.3333 | | |
| 4 | M | C | 24 | 1 | 185 | | 1035.9784 | | |
| 4 | M | C | 120 | 0 | 19 | | 739.3684 | | |
| 4 | M | C | 120 | 1 | 181 | | 985.8122 | | |
| 7 | F | C | 0 | 0 | 16 | | 809.1875 | |
| 7 | F | C | 0 | 1 | 164 | | 997.7500 | |
| 7 | M | C | 0 | 0 | 17 | | 971.0000 | |
| 7 | M | C | 0 | 1 | 293 | | 987.4096 | |
| 9 | F | C | 1 | 0 | 110 | | 880.0545 | |
| 9 | F | C | 1 | 1 | 840 | | 984.1226 | |
| 9 | F | C | 24 | 0 | 14 | | 1002.2857 | |
| 9 | F | C | 24 | 1 | 186 | | 968.3871 | |
| 13 | F | C | 60 | 0 | 59 | | 725.0169 | |
| 13 | F | C | 60 | 1 | 541 | | 989.2440 | |
| 13 | M | C | 60 | 0 | 49 | | 642.3673 | |
| 13 | M | C | 60 | 1 | 565 | | 991.3558 | |
| 14 | F | C | 1 | 0 | 12 | | 919.7500 | |
| 14 | F | C | 1 | 1 | 188 | | 981.3989 | |
| 14 | M | C | 1 | 0 | 11 | | 940.1818 | |
| 14 | M | C | 1 | 1 | 189 | | 1014.4339 | |

In previous analysis with KM survival, JM4 (excluding 120 and 300 fractions, excluding males). Not sure why JM14 is suddenly giving me issues… first exclude JM4 males and see what happens.

graphs <- cox.censored(stratified.by = c("sex", "expt", "fractions"), df = subset(all.info, total\_dose == 0 & !(expt == 9 & sex == "M") & expt != 8 & fractions != 300 & !(expt == 4 & sex == "M")))

Surv(age, status) ~ sex + expt + fractions

<environment: 0x11d131a00>

Call:

coxph(formula = f, data = df)

n= 4135, number of events= 3771

coef exp(coef) se(coef) z Pr(>|z|)

sexM -0.03631 0.96434 0.04199 -0.865 0.3871

expt4 -0.12151 0.88558 0.11027 -1.102 0.2705

expt7 -0.20826 0.81200 0.10745 -1.938 0.0526 .

expt8 NA NA 0.00000 NA NA

expt9 -0.12416 0.88324 0.06725 -1.846 0.0649 .

expt13 -0.20455 0.81502 0.10127 -2.020 0.0434 \*

expt14 -0.12480 0.88267 0.07378 -1.692 0.0907 .

fractions1 -0.11176 0.89426 0.08109 -1.378 0.1681

fractions24 NA NA 0.00000 NA NA

fractions60 NA NA 0.00000 NA NA

fractions120 NA NA 0.00000 NA NA

Concordance= 0.515 (se = 0.005 )

Rsquare= 0.002 (max possible= 1 )

Likelihood ratio test= 9.24 on 7 df, p=0.2358

Wald test = 9.37 on 7 df, p=0.2274

Score (logrank) test = 9.38 on 7 df, p=0.2268

rho chisq p

sexM 0.01114 0.4671 0.4943

expt4 0.01529 0.8872 0.3463

expt7 0.00857 0.2767 0.5989

expt9 0.02385 2.1790 0.1399

expt13 0.01786 1.2041 0.2725

expt14 0.02793 2.9515 0.0858

fractions1 0.00222 0.0186 0.8917

fractions24 NA NaN NaN

fractions60 NA NaN NaN

fractions120 NA NaN NaN

GLOBAL NA 4.4325 1.0000

The overall model isn’t significant, which is all I checked previously with KM. Now I can see how individual levels of fractions compare to the reference and check for significance.

First – figure out issue with NA values… when I run the regression with expt listed first, I get NA’s for 60 and 120 fractions. When I run the regression with fractions first, I get NA for expt 7 and 13. 120 fractions comes from expt 4 and 60 fractions comes from expt 13.

Back to the basics…

Maybe change the reference experiment?

Last analysis these were my results:

Include the following data-

JM3

JM4 (excluding 120 and 300 fractions, excluding males)

JM7

JM8\* only helpful to compare dose rates

JM9 (excluding males)

JM13

JM14

Maybe find a way to subtract out changes in controls?

Maybe eliminate a lot more data?

Maybe add experiment to my list of covariates to control for them?

Maybe graphing things would help explain them better?

Why can’t I see values for everything?

START OVER:

Use all data after filtering as shown in tables at the beginning (for example, JM 12 is included).

graphs <- cox.censored(stratified.by = c("sex", "expt", "fractions"), df = subset(all.info, total\_dose == 0))

Surv(age, status) ~ sex + expt + fractions

<environment: 0x11ba0fed8>

Call:

coxph(formula = f, data = df)

n= 5140, number of events= 4708

coef exp(coef) se(coef) z Pr(>|z|)

sexM -1.210e-01 8.861e-01 3.357e-02 -3.604 0.000314 \*\*\*

expt4 -2.622e-01 7.693e-01 1.036e-01 -2.532 0.011338 \*

expt7 -6.660e-01 5.138e-01 1.377e-01 -4.835 1.33e-06 \*\*\*

expt8 4.662e-01 1.594e+00 1.008e+00 0.462 0.643817

expt9 -1.948e-01 8.230e-01 6.225e-02 -3.129 0.001753 \*\*

expt12 -8.673e-01 4.201e-01 1.600e-01 -5.421 5.92e-08 \*\*\*

expt13 -6.715e-01 5.109e-01 1.334e-01 -5.034 4.81e-07 \*\*\*

expt14 -1.266e-01 8.811e-01 7.373e-02 -1.716 0.086092 .

fractions1 -5.745e-01 5.630e-01 1.186e-01 -4.845 1.27e-06 \*\*\*

fractions24 -4.362e-01 6.465e-01 8.905e-02 -4.899 9.64e-07 \*\*\*

fractions60 NA NA 0.000e+00 NA NA

fractions120 -4.080e-01 6.650e-01 1.068e-01 -3.822 0.000133 \*\*\*

fractions300 NA NA 0.000e+00 NA NA

Concordance= 0.537 (se = 0.005 )

Rsquare= 0.03 (max possible= 1 )

Likelihood ratio test= 156.8 on 75 df, p=1.006e-07

Wald test = 207.9 on 75 df, p=1.998e-14

Score (logrank) test = 326.8 on 75 df, p=0

rho chisq p

sexM 1.58e-02 1.14e+00 0.2867

expt4 1.58e-02 1.17e+00 0.2792

expt7 8.20e-03 3.16e-01 0.5738

expt8 1.30e-02 7.97e-01 0.3719

expt9 2.36e-02 2.62e+00 0.1057

expt12 7.11e-03 2.38e-01 0.6257

expt13 1.48e-02 1.04e+00 0.3089

expt14 2.58e-02 3.14e+00 0.0765

fractions1 4.28e-03 8.58e-02 0.7696

fractions15 NA NaN NaN

fractions24 4.49e-03 9.42e-02 0.7589

fractions60 NA NaN NaN

fractions120 -2.05e-02 1.98e+00 0.1592

fractions300 NA NaN NaN

Anything with NA had warning:

In coxph(f, data = df) :

X matrix deemed to be singular; variable 10 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 29 30 31 32 33 34 35 38 40 42 43 44 45 46 47 48 49 50 54 56 63 68 69 72 73 79 83 84 92 94 99 106 110 111 119 121 122 126 127 129 131 134

sexM 1.58e-02 1.14e+00 0.2867

expt4 1.58e-02 1.17e+00 0.2792

expt7 8.20e-03 3.16e-01 0.5738

expt8 1.30e-02 7.97e-01 0.3719

expt9 2.36e-02 2.62e+00 0.1057

expt12 7.11e-03 2.38e-01 0.6257

expt13 1.48e-02 1.04e+00 0.3089

expt14 2.58e-02 3.14e+00 0.0765

fractions1 4.28e-03 8.58e-02 0.7696

fractions15 NA NaN NaN

fractions24 4.49e-03 9.42e-02 0.7589

fractions30 NA NaN NaN

fractions31 NA NaN NaN

fractions33 NA NaN NaN

fractions37 NA NaN NaN

fractions38 NA NaN NaN

fractions39 NA NaN NaN

fractions49 NA NaN NaN

fractions50 NA NaN NaN

fractions51 NA NaN NaN

fractions53 NA NaN NaN

fractions57 NA NaN NaN

fractions58 NA NaN NaN

fractions60 NA NaN NaN

fractions61 NA NaN NaN

fractions62 NA NaN NaN

fractions63 NA NaN NaN

fractions64 -7.68e-03 2.78e-01 0.5977

fractions66 NA NaN NaN

fractions67 NA NaN NaN

fractions68 NA NaN NaN

fractions69 NA NaN NaN

fractions70 NA NaN NaN

fractions71 NA NaN NaN

fractions72 NA NaN NaN

fractions73 -7.06e-03 2.35e-01 0.6278

fractions74 -7.07e-03 2.35e-01 0.6277

fractions75 NA NaN NaN

fractions76 -7.33e-03 2.53e-01 0.6149

fractions77 NA NaN NaN

fractions78 -7.18e-03 2.43e-01 0.6222

fractions79 NA NaN NaN

fractions80 NA NaN NaN

fractions81 NA NaN NaN

fractions82 NA NaN NaN

fractions83 NA NaN NaN

fractions84 NA NaN NaN

fractions85 NA NaN NaN

fractions86 NA NaN NaN

fractions87 NA NaN NaN

fractions89 -6.12e-03 1.76e-01 0.6747

fractions90 1.12e-02 2.91e-07 0.9996

fractions91 -6.44e-03 1.96e-01 0.6583

fractions92 NA NaN NaN

fractions93 -6.04e-03 1.72e-01 0.6783

fractions94 NA NaN NaN

fractions95 -6.79e-03 2.17e-01 0.6410

fractions96 -6.80e-03 2.18e-01 0.6405

fractions97 -6.97e-03 2.29e-01 0.6321

fractions98 -6.26e-03 1.84e-01 0.6676

fractions99 9.76e-03 3.63e-07 0.9995

fractions100 -5.84e-03 1.61e-01 0.6884

fractions101 NA NaN NaN

fractions102 9.10e-03 3.77e-07 0.9995

fractions103 -5.39e-03 1.37e-01 0.7116

fractions104 -4.41e-03 9.14e-02 0.7624

fractions105 -4.36e-03 8.94e-02 0.7649

fractions106 NA NaN NaN

fractions107 NA NaN NaN

fractions108 -3.51e-03 5.79e-02 0.8098

fractions109 -3.64e-03 6.25e-02 0.8026

fractions110 NA NaN NaN

fractions111 NA NaN NaN

fractions112 -3.34e-03 5.26e-02 0.8185

fractions113 -2.99e-03 4.22e-02 0.8372

fractions114 -2.70e-03 3.44e-02 0.8528

fractions115 -3.26e-03 5.00e-02 0.8230

fractions116 4.21e-03 1.75e-07 0.9997

fractions117 NA NaN NaN

fractions118 -2.29e-03 2.48e-02 0.8750

fractions119 -2.53e-03 3.01e-02 0.8623

fractions120 -2.05e-02 1.98e+00 0.1592

fractions121 NA NaN NaN

fractions122 NA NaN NaN

fractions123 -9.21e-04 3.99e-03 0.9496

fractions124 -1.04e-03 5.08e-03 0.9432

fractions125 -5.54e-04 1.45e-03 0.9697

fractions126 -4.50e-04 9.54e-04 0.9754

fractions127 -3.94e-05 7.30e-06 0.9978

fractions128 -3.94e-05 7.30e-06 0.9978

fractions129 1.23e-04 7.15e-05 0.9933

fractions130 NA NaN NaN

fractions131 7.85e-04 2.90e-03 0.9570

fractions132 NA NaN NaN

fractions133 6.35e-04 1.90e-03 0.9653

fractions134 6.79e-04 2.17e-03 0.9629

fractions135 1.06e-03 5.32e-03 0.9419

fractions136 1.47e-03 1.02e-02 0.9197

fractions137 NA NaN NaN

fractions138 1.73e-03 1.41e-02 0.9056

fractions139 1.74e-03 1.42e-02 0.9051

fractions140 1.80e-03 1.52e-02 0.9018

fractions141 1.82e-03 1.56e-02 0.9007

fractions142 2.70e-03 3.43e-02 0.8531

fractions143 2.52e-03 2.99e-02 0.8627

fractions144 NA NaN NaN

fractions145 3.01e-03 4.27e-02 0.8363

fractions146 -1.07e-02 4.15e-06 0.9984

fractions147 2.67e-03 3.35e-02 0.8548

fractions148 NA NaN NaN

fractions149 NA NaN NaN

fractions150 3.06e-03 4.42e-02 0.8335

fractions151 2.23e-03 2.33e-02 0.8787

fractions152 3.02e-03 4.29e-02 0.8359

fractions153 2.82e-03 3.74e-02 0.8467

fractions154 2.46e-03 2.84e-02 0.8661

fractions155 2.43e-03 2.79e-02 0.8674

fractions156 2.37e-03 2.65e-02 0.8707

fractions157 NA NaN NaN

fractions158 2.30e-03 2.49e-02 0.8747

fractions159 NA NaN NaN

fractions160 NA NaN NaN

fractions161 2.25e-03 2.38e-02 0.8773

fractions163 2.10e-03 2.08e-02 0.8852

fractions165 1.69e-03 1.34e-02 0.9077

fractions167 NA NaN NaN

fractions168 NA NaN NaN

fractions169 1.33e-03 8.31e-03 0.9273

fractions171 NA NaN NaN

fractions173 1.17e-03 6.44e-03 0.9360

fractions175 NA NaN NaN

fractions177 2.51e-04 2.95e-04 0.9863

fractions183 -1.11e-03 5.76e-03 0.9395

fractions300 NA NaN NaN

Figured out my issue!!!

<https://stackoverflow.com/questions/20977401/coxph-x-matrix-deemed-to-be-singular>

The error “In coxph(f, data = df) :

X matrix deemed to be singular;” and the NA values are because of “perfect qualifiers” meaning I was using too many variables and didn’t have enough diverse data. This will not be an issue during my other analysis work because I will never use expt as a variable. I’m only doing that now because I’m determining which experiments can be compared to one another.

The main goal is to see which experiments could be compared, so there’s no need to also look at fractions at the same time. The mail goal for fractions is to just look at all fractions and if the larger amounts caused mice to die earlier, but that doesn’t need to consider experiment. Fine to keep them separated out. Still include sex because that always plays a role/should play a role and needs to be controlled for.

To do:  
Sex, fractions, expt 🡪 do all individually

Sex/fractions and sex/expt combos 🡪 do both pairs

Sex, use all data:

graphs <- cox.censored(stratified.by = c("sex"), df = subset(all.info, total\_dose == 0))

Surv(age, status) ~ sex

<environment: 0x1121410a0>

Call:

coxph(formula = f, data = df)

n= 5140, number of events= 4708

coef exp(coef) se(coef) z Pr(>|z|)

sexM -0.08695 0.91672 0.02933 -2.964 0.00303 \*\*

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

exp(coef) exp(-coef) lower .95 upper .95

sexM 0.9167 1.091 0.8655 0.971

Concordance= 0.511 (se = 0.004 )

Rsquare= 0.002 (max possible= 1 )

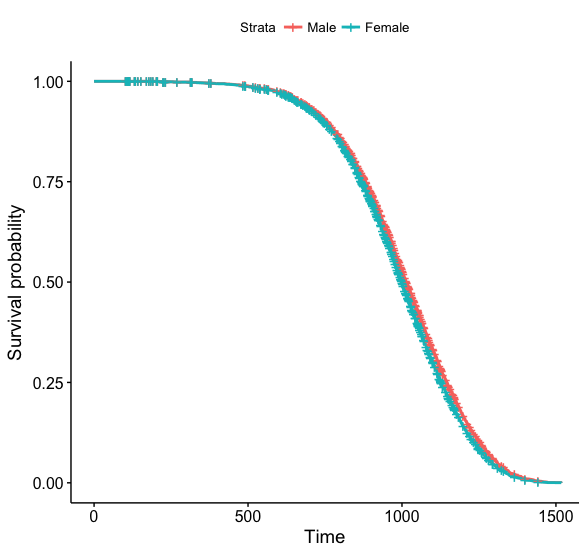
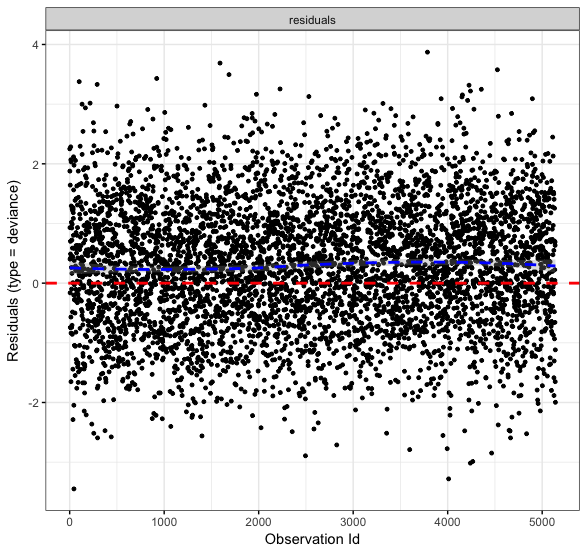
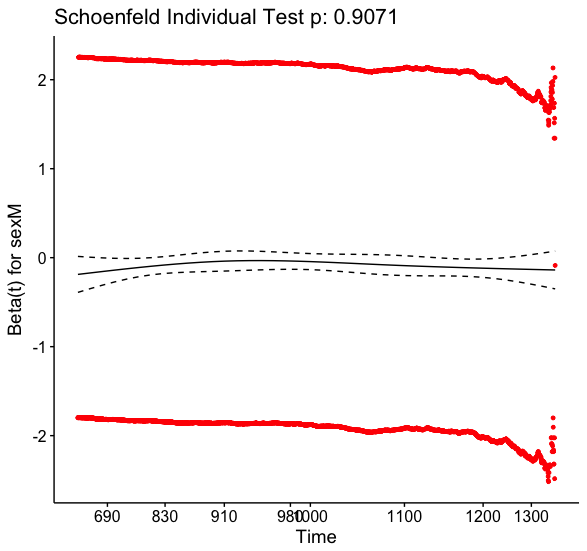
Likelihood ratio test= 8.81 on 1 df, p=0.002997

Wald test = 8.79 on 1 df, p=0.003033

Score (logrank) test = 8.79 on 1 df, p=0.003023

rho chisq p

sexM -0.0017 0.0136 0.907

Expt use all data:

graphs <- cox.censored(stratified.by = c("expt"), df = subset(all.info, total\_dose == 0))

Surv(age, status) ~ expt

<environment: 0x132070630>

Call:

coxph(formula = f, data = df)

n= 5140, number of events= 4708

coef exp(coef) se(coef) z Pr(>|z|)

expt4 -0.05495 0.94653 0.06158 -0.892 0.3722

expt7 -0.09736 0.90723 0.07045 -1.382 0.1670

expt8 -0.10035 0.90452 0.11395 -0.881 0.3785

expt9 -0.12957 0.87847 0.06000 -2.160 0.0308 \*

expt12 -0.34169 0.71057 0.10825 -3.157 0.0016 \*\*

expt13 -0.08983 0.91408 0.06065 -1.481 0.1386

expt14 -0.12067 0.88633 0.07372 -1.637 0.1017

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

exp(coef) exp(-coef) lower .95 upper .95

expt4 0.9465 1.056 0.8389 1.0680

expt7 0.9072 1.102 0.7902 1.0416

expt8 0.9045 1.106 0.7235 1.1309

expt9 0.8785 1.138 0.7810 0.9881

expt12 0.7106 1.407 0.5747 0.8785

expt13 0.9141 1.094 0.8116 1.0295

expt14 0.8863 1.128 0.7671 1.0241

Concordance= 0.515 (se = 0.005 )

Rsquare= 0.003 (max possible= 1 )

Likelihood ratio test= 13.83 on 7 df, p=0.05422

Wald test = 13.36 on 7 df, p=0.06387

Score (logrank) test = 13.4 on 7 df, p=0.06286

rho chisq p

expt4 0.01260 0.748 0.3873

expt7 0.01049 0.518 0.4715

expt8 0.01768 1.472 0.2250

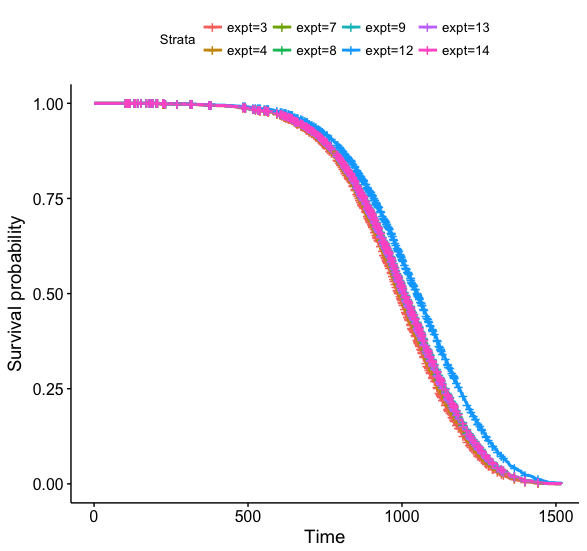
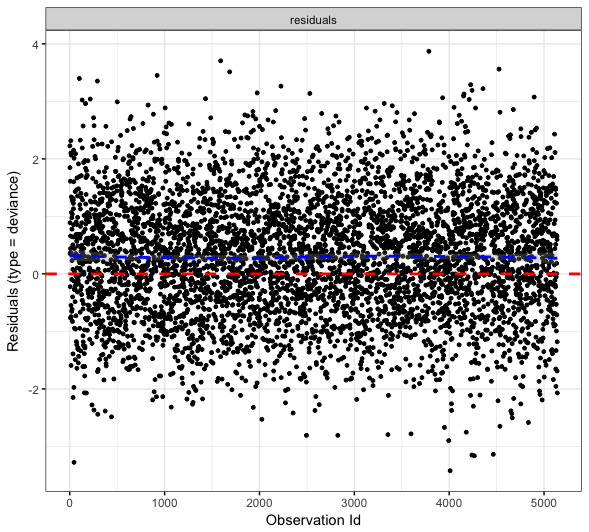
expt9 0.02015 1.916 0.1663

expt12 0.00889 0.371 0.5424

expt13 0.02424 2.771 0.0960

expt14 0.02594 3.174 0.0748

GLOBAL NA 5.169 0.6394

fractions use all data:

graphs <- cox.censored(stratified.by = "fractions", df = subset(all.info, total\_dose == 0 & expt != 8))

Surv(age, status) ~ fractions

<environment: 0x130dc0e38>

Call:

coxph(formula = f, data = df)

n= 5030, number of events= 4610

coef exp(coef) se(coef) z Pr(>|z|)

fractions1 0.034291 1.034886 0.048229 0.711 0.477

fractions24 0.075440 1.078358 0.054686 1.380 0.168

fractions60 0.060852 1.062742 0.051641 1.178 0.239

fractions120 0.002765 1.002769 0.085438 0.032 0.974

fractions300 0.405402 1.499905 0.087240 4.647 3.37e-06 \*\*\*

Concordance= 0.509 (se = 0.005 )

Rsquare= 0.004 (max possible= 1 )

Likelihood ratio test= 22.14 on 5 df, p=0.0004931

Wald test = 24.33 on 5 df, p=0.0001879

Score (logrank) test = 24.57 on 5 df, p=0.0001686

rho chisq p

fractions1 0.01043 0.502 0.479

fractions15 NA NaN NaN

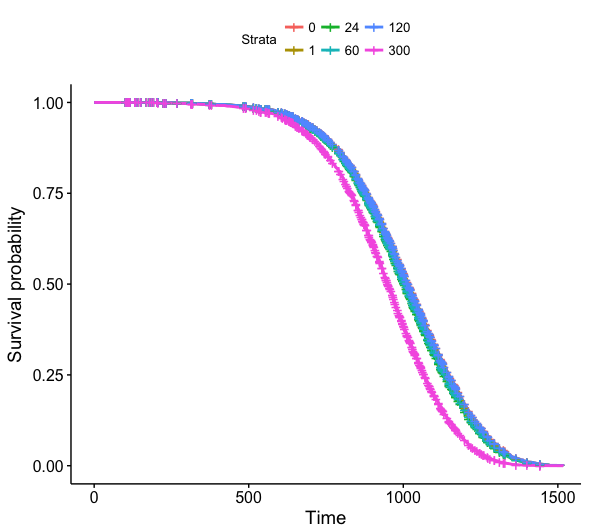
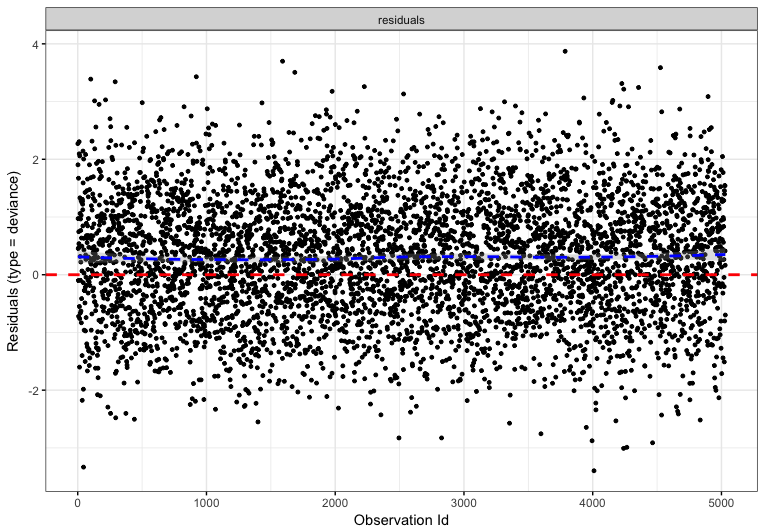
fractions24 0.01197 0.660 0.417

fractions60 0.01633 1.231 0.267

fractions120 -0.01635 1.233 0.267

fractions300 0.00842 0.328 0.567

GLOBAL NA 4.371 1.000

sex + expt all data

> graphs <- cox.censored(stratified.by = c("sex", "expt"), df = subset(all.info, total\_dose == 0))

Surv(age, status) ~ sex + expt

<environment: 0x157960c68>

Call:

coxph(formula = f, data = df)

n= 5140, number of events= 4708

coef exp(coef) se(coef) z Pr(>|z|)

sexM -0.10237 0.90269 0.03146 -3.254 0.00114 \*\*

expt4 -0.05415 0.94729 0.06158 -0.879 0.37916

expt7 -0.09165 0.91242 0.07047 -1.301 0.19339

expt8 -0.09236 0.91177 0.11397 -0.810 0.41771

expt9 -0.16758 0.84571 0.06109 -2.743 0.00608 \*\*

expt12 -0.29698 0.74306 0.10917 -2.720 0.00652 \*\*

expt13 -0.09562 0.90881 0.06067 -1.576 0.11503

expt14 -0.12495 0.88254 0.07373 -1.695 0.09012 .

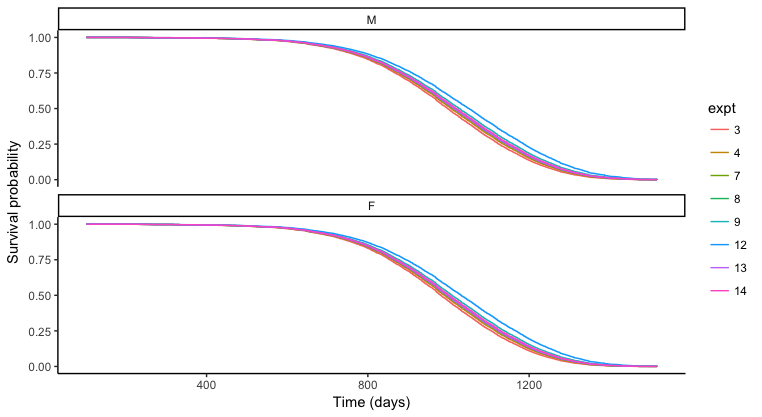
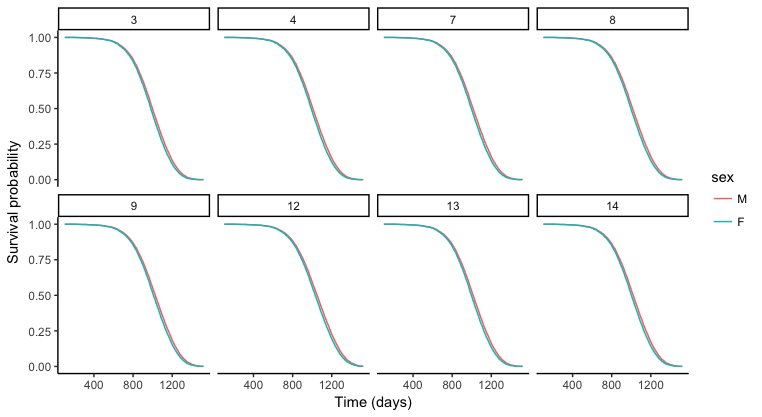
Concordance= 0.52 (se = 0.005 )

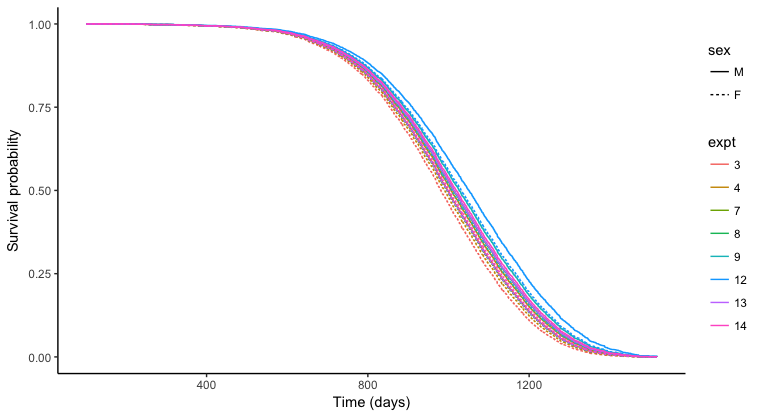
Rsquare= 0.005 (max possible= 1 )

Likelihood ratio test= 24.44 on 8 df, p=0.001931

Wald test = 24.06 on 8 df, p=0.002237

Score (logrank) test = 24.12 on 8 df, p=0.002187





rho chisq p

sexM 0.00308 0.044 0.8339

expt4 0.01246 0.731 0.3927

expt7 0.00939 0.415 0.5194

expt8 0.01857 1.626 0.2022

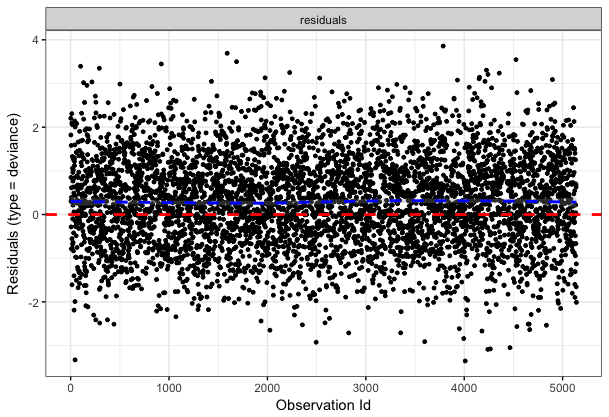
expt9 0.02051 1.980 0.1594

expt12 0.00793 0.295 0.5872

expt13 0.02368 2.646 0.1038

expt14 0.02546 3.057 0.0804

GLOBAL NA 5.228 0.7329



sex + fractions all data

graphs <- cox.censored(stratified.by = c("sex", "fractions"), df = subset(all.info, total\_dose == 0 & expt !=8))

Surv(age, status) ~ sex + fractions

<environment: 0x1611d31c0>

Call:

coxph(formula = f, data = df)

n= 5030, number of events= 4610

coef exp(coef) se(coef) z Pr(>|z|)

sexM -0.113778 0.892456 0.032564 -3.494 0.000476 \*\*\*

fractions1 -0.006899 0.993124 0.049656 -0.139 0.889494

fractions24 0.027552 1.027935 0.056375 0.489 0.625040

fractions60 0.037891 1.038618 0.052067 0.728 0.466771

fractions120 0.036491 1.037165 0.086025 0.424 0.671424

fractions300 0.440320 1.553204 0.087857 5.012 5.39e-07 \*\*\*

Concordance= 0.521 (se = 0.005 )

Rsquare= 0.007 (max possible= 1 )

Likelihood ratio test= 34.42 on 6 df, p=5.574e-06

Wald test = 36.39 on 6 df, p=2.311e-06

Score (logrank) test = 36.68 on 6 df, p=2.035e-06

rho chisq p

sexM 0.01369 0.843 0.359

fractions1 0.01432 0.936 0.333

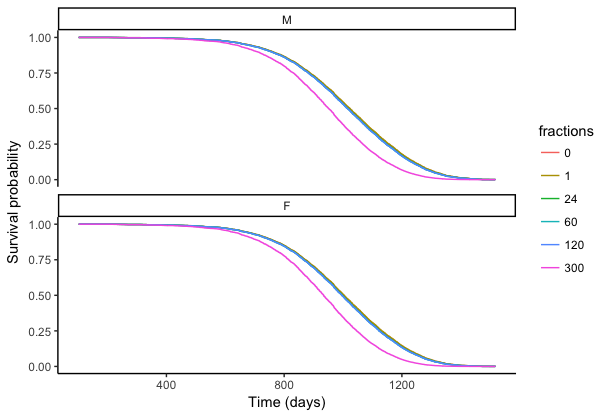
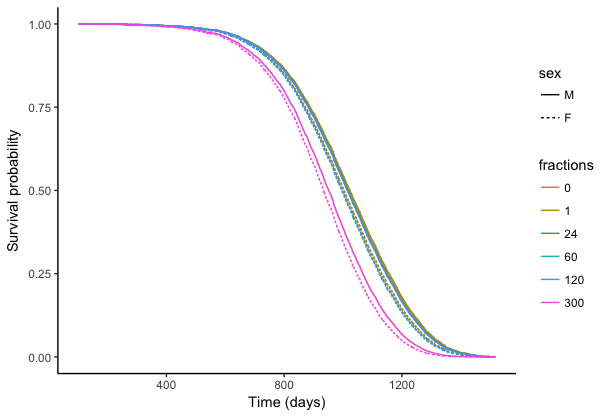
fractions24 0.01642 1.220 0.269

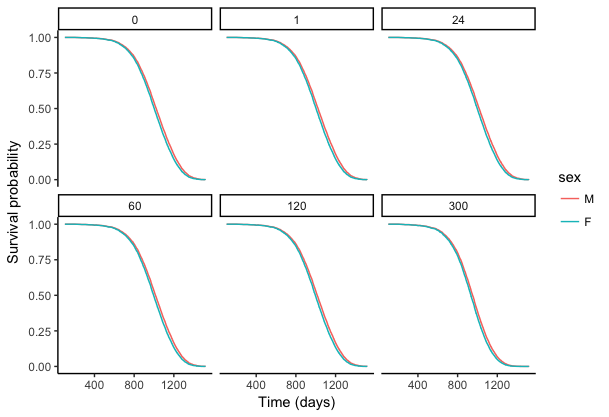
fractions60 0.01807 1.506 0.220

fractions120 -0.01790 1.480 0.224

fractions300 0.00691 0.221 0.638

GLOBAL NA 5.402 1.000





\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* based on round 1 – remove JM12, JM9 males, and 300 fractions \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

sex + fractions

Call:

coxph(formula = Surv(age, status) ~ sex + fractions, data = subset(all.info.no8,

total\_dose == 0))

n= 4535, number of events= 4137

coef exp(coef) se(coef) z Pr(>|z|)

sexM -0.0745524 0.9281589 0.0346694 -2.150 0.0315 \*

fractions1 -0.0151169 0.9849967 0.0546792 -0.276 0.7822

fractions24 -0.0026810 0.9973226 0.0595978 -0.045 0.9641

fractions60 -0.0008275 0.9991728 0.0557742 -0.015 0.9882

fractions120 -0.0216311 0.9786011 0.0889683 -0.243 0.8079

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

exp(coef) exp(-coef) lower .95 upper .95

sexM 0.9282 1.077 0.8672 0.9934

fractions1 0.9850 1.015 0.8849 1.0964

fractions24 0.9973 1.003 0.8874 1.1209

fractions60 0.9992 1.001 0.8957 1.1146

fractions120 0.9786 1.022 0.8220 1.1650

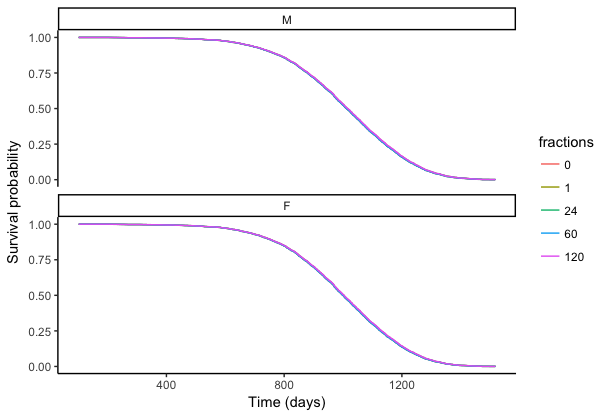
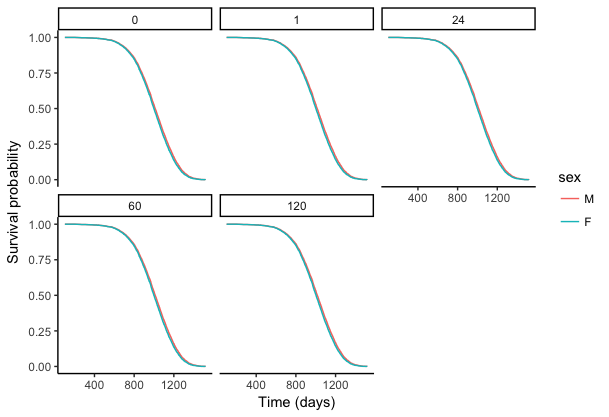
Concordance= 0.51 (se = 0.005 )

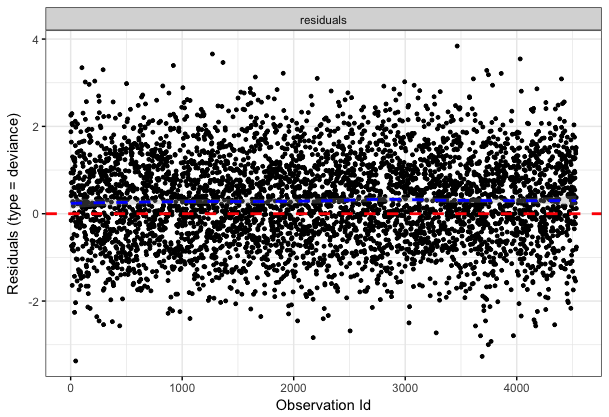
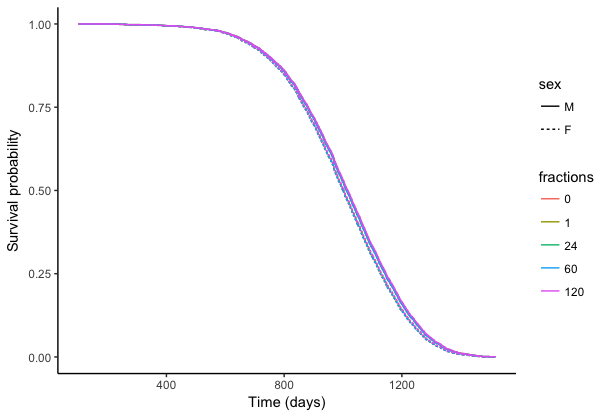
Rsquare= 0.001 (max possible= 1 )

Likelihood ratio test= 5.49 on 5 df, p=0.3587

Wald test = 5.46 on 5 df, p=0.3619

Score (logrank) test = 5.47 on 5 df, p=0.3616





rho chisq p

sexM 0.0130 0.688 0.407

fractions1 0.0111 0.502 0.479

fractions24 0.0136 0.754 0.385

fractions60 0.0155 0.990 0.320

fractions120 -0.0192 1.536 0.215

GLOBAL NA 4.767 0.445

sex + expt

coxph(formula = Surv(age, status) ~ sex + expt, data = subset(all.info,

total\_dose == 0))

n= 4645, number of events= 4235

coef exp(coef) se(coef) z Pr(>|z|)

sexM -0.10143 0.90355 0.03550 -2.857 0.00427 \*\*

expt4 -0.11681 0.88975 0.06335 -1.844 0.06518 .

expt7 -0.09319 0.91102 0.07048 -1.322 0.18609

expt8 -0.09421 0.91009 0.11398 -0.826 0.40852

expt9 -0.14198 0.86764 0.06421 -2.211 0.02703 \*

expt13 -0.09703 0.90753 0.06068 -1.599 0.10985

expt14 -0.12732 0.88046 0.07374 -1.727 0.08423 .

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

exp(coef) exp(-coef) lower .95 upper .95

sexM 0.9035 1.107 0.8428 0.9687

expt4 0.8898 1.124 0.7859 1.0074

expt7 0.9110 1.098 0.7935 1.0460

expt8 0.9101 1.099 0.7279 1.1379

expt9 0.8676 1.153 0.7650 0.9840

expt13 0.9075 1.102 0.8058 1.0222

expt14 0.8805 1.136 0.7620 1.0174

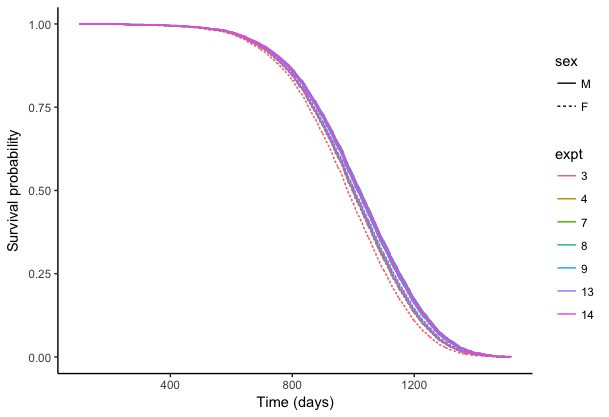
Concordance= 0.516 (se = 0.005 )

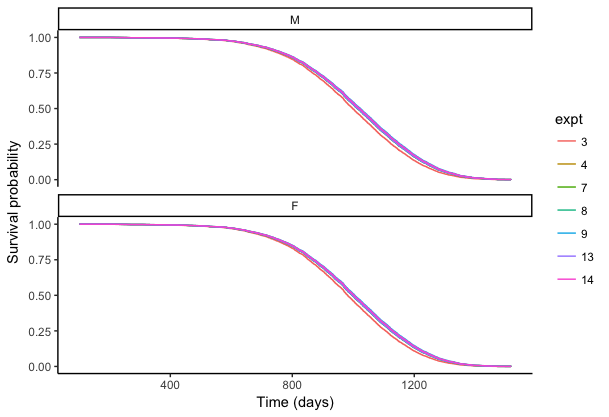
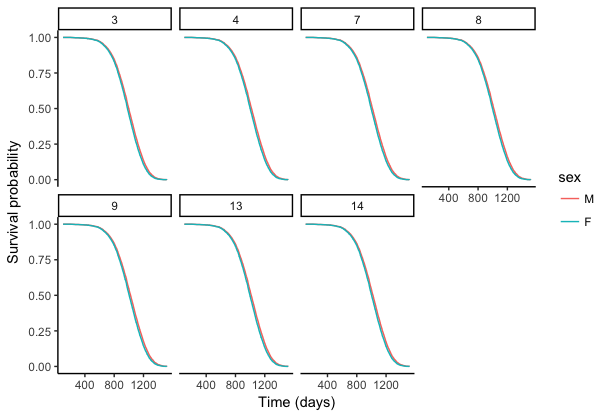
Rsquare= 0.003 (max possible= 1 )

Likelihood ratio test= 11.84 on 7 df, p=0.1058

Wald test = 11.92 on 7 df, p=0.1031

Score (logrank) test = 11.93 on 7 df, p=0.1029



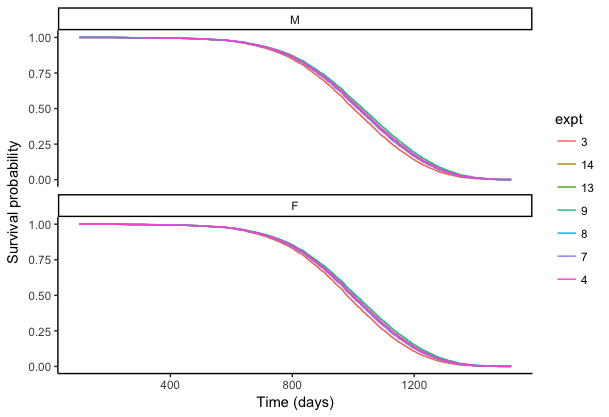


Expt 9 is still significant and expt 4 and 14 are very close to being significant, but the entire model is NOT significant. Graphing it, I saw that expt 3 appeared to be more of an outlier than any other experiment. If I change the reference level for expt, I can see that there isn’t any significance except when expt 3 is the reference level. When expt 9 is the reference, expt 3 shows signifance. When expt 14 is the reference, it shows expt 3 is very close to being significant (makes sense because the reverse was true too).

Expt3 doesn’t have any fractionation, but I do like the total doses used… Expt 3 has 1593 controls, while expt 9 has 2400 controls. Expt 9 has no fractionation, very low doses, but mostly neutron treatments.

Went back and checked for fractions… changing the reference level doesn’t impact significance.

What if I add back in the expt 9 males? Graph didn’t show that they actually looked bad earlier. Now I see that the overall model is significant (could be due to sex being significant), but that expt 9 is only an issue with 3 is the reference. Graph below…



Still looks like expt3 is the strange one and not expt 9.

Check that if sex is removed the overall model is no longer significant? Yes, holds true.

What do I do???????? Because expt 3 looks to be the one hanging out and Tanja said that as time went on the conditions were optimized, it makes sense that expt3 would be dying earlier and more of an outlier. Remove expt 3 and see what happens.

coxph(formula = Surv(age, status) ~ expt, data = subset(all.info,

total\_dose == 0 & expt != 3))

n= 4460, number of events= 4062

coef exp(coef) se(coef) z Pr(>|z|)

expt3 NA NA 0.000000 NA NA

expt14 0.008874 1.008913 0.058986 0.150 0.880

expt13 0.039877 1.040682 0.041583 0.959 0.338

expt8 0.029186 1.029616 0.105043 0.278 0.781

expt7 0.031123 1.031612 0.054899 0.567 0.571

expt4 0.019218 1.019404 0.045409 0.423 0.672

exp(coef) exp(-coef) lower .95 upper .95

expt3 NA NA NA NA

expt14 1.009 0.9912 0.8988 1.133

expt13 1.041 0.9609 0.9592 1.129

expt8 1.030 0.9712 0.8380 1.265

expt7 1.032 0.9694 0.9264 1.149

expt4 1.019 0.9810 0.9326 1.114

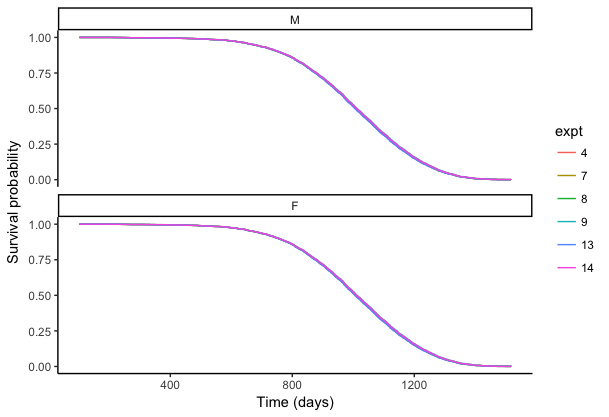
Concordance= 0.504 (se = 0.005 )

Rsquare= 0 (max possible= 1 )

Likelihood ratio test= 1.03 on 5 df, p=0.9598

Wald test = 1.03 on 5 df, p=0.9598

Score (logrank) test = 1.03 on 5 df, p=0.9598



coxph(formula = Surv(age, status) ~ sex + expt, data = subset(all.info,

total\_dose == 0 & expt != 3))

n= 4460, number of events= 4062

coef exp(coef) se(coef) z Pr(>|z|)

sexM -0.126501 0.881173 0.033755 -3.748 0.000179 \*\*\*

expt9 -0.050301 0.950943 0.059973 -0.839 0.401625

expt13 0.029130 1.029559 0.059664 0.488 0.625381

expt8 0.035645 1.036288 0.113480 0.314 0.753441

expt7 0.034540 1.035143 0.069688 0.496 0.620153

expt4 0.008696 1.008734 0.062396 0.139 0.889158

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

exp(coef) exp(-coef) lower .95 upper .95

sexM 0.8812 1.1349 0.8248 0.9414

expt9 0.9509 1.0516 0.8455 1.0696

expt13 1.0296 0.9713 0.9159 1.1573

expt8 1.0363 0.9650 0.8296 1.2944

expt7 1.0351 0.9661 0.9030 1.1866

expt4 1.0087 0.9913 0.8926 1.1400

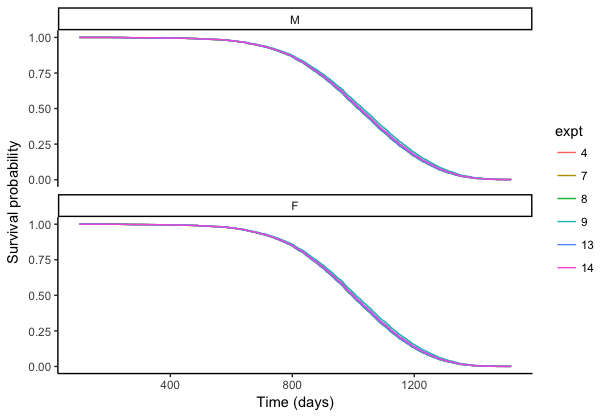
Concordance= 0.518 (se = 0.005 )

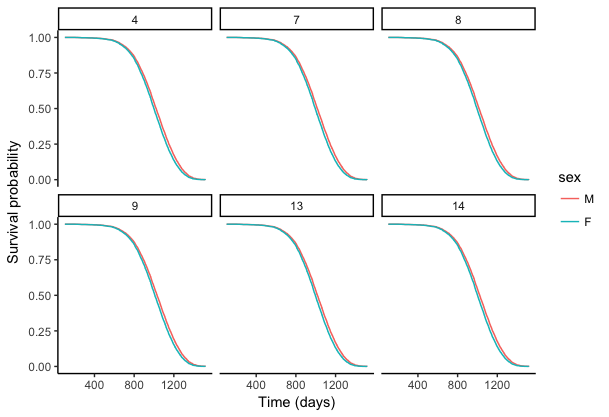
Rsquare= 0.003 (max possible= 1 )

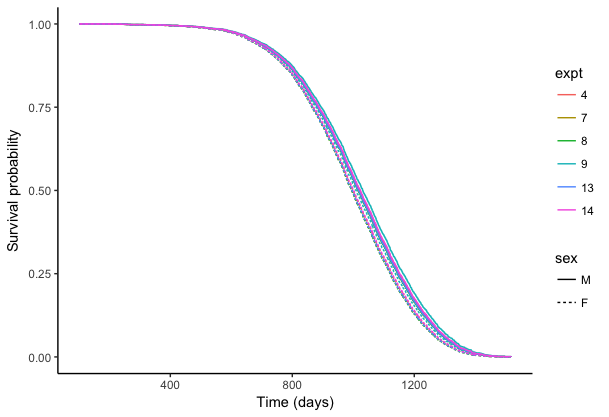
Likelihood ratio test= 15.14 on 6 df, p=0.01919

Wald test = 15.12 on 6 df, p=0.01936

Score (logrank) test = 15.13 on 6 df, p=0.01929







Data that we will now be missing by excluding experiment 3….

| **sex** | | **radn** | | **fractions** | | **status** | | **total\_dose** | | **n** | **mean\_age** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  | |  |  |  | |  |  |
| **1** | F | C | | 1 | 0 | 0.00 | | 14 | 798.6429 |
| **2** | F | C | | 1 | 1 | 0.00 | | 171 | 942.8538 |
| **3** | F | G | | 1 | 0 | 86.31 | | 11 | 897.0000 |
| **4** | F | G | | 1 | 0 | 137.14 | | 35 | 450.0000 |
| **5** | F | G | | 1 | 0 | 197.55 | | 37 | 450.0000 |
| **6** | F | G | | 1 | 0 | 399.90 | | 6 | 739.8333 |
| **7** | F | G | | 1 | 0 | 545.70 | | 30 | 445.8667 |
| **8** | F | G | | 1 | 1 | 86.31 | | 189 | 938.2487 |
| **9** | F | G | | 1 | 1 | 137.14 | | 7 | 767.2857 |
| **10** | F | G | | 1 | 1 | 197.55 | | 6 | 430.6667 |
| **11** | F | G | | 1 | 1 | 399.90 | | 54 | 826.3519 |
| **12** | F | G | | 1 | 1 | 545.70 | | 74 | 731.3243 |
| **13** | M | C | | 1 | 0 | 0.00 | | 9 | 948.7778 |
| **14** | M | C | | 1 | 1 | 0.00 | | 191 | 987.0576 |
| **15** | M | G | | 1 | 0 | 86.31 | | 11 | 789.0909 |
| **16** | M | G | | 1 | 0 | 137.14 | | 10 | 818.8000 |
| **17** | M | G | | 1 | 0 | 197.55 | | 5 | 784.0000 |
| **18** | M | G | | 1 | 0 | 399.90 | | 3 | 544.6667 |
| **19** | M | G | | 1 | 0 | 545.70 | | 2 | 746.5000 |
| **20** | M | G | | 1 | 1 | 86.31 | | 189 | 976.1693 |
| **21** | M | G | | 1 | 1 | 137.14 | | 150 | 947.1800 |
| **22** | M | G | | 1 | 1 | 197.55 | | 155 | 918.4387 |
| **23** | M | G | | 1 | 1 | 399.90 | | 117 | 864.0342 |
| **24** | M | G | | 1 | 1 | 545.70 | | 118 | 758.6780 |

Change reference expt and double check that there’s never any significance…