BSTA 662: Survival Analysis Final Report

Exploring Factors Influencing Infant Pneumonia Hospitalization

Jamie Calma

Spring 2024

Abstract:

Pneumonia remains a significant cause of death, particularly among children in the world. In this study, we investigated the impact of infant dietary choices, specifically the timing of weaning off breast milk, on infants' susceptibility to pneumonia-related hospitalizations. Leveraging 3,470 mothers' interview data from the National Longitudinal Survey of Youth, our analysis revealed that extending breastfeeding by one month was associated with a 19% reduction in the risk of infant pneumonia hospitalization. Maternal age, number of siblings, and maternal smoking habits during pregnancy also emerged as significant factors to infant pneumonia-related hospitalizations. These findings highlight the importance of maternal health behaviors and infant feeding practices in mitigating infant pneumonia hospitalization risk. Overall, the study highlights the complex interplay of socio-demographic and environmental factors in infants and calls for targeted interventions to improve health outcomes.

Introduction:

Pneumonia, a respiratory infection, stands as a significant cause of mortality, particularly among children, claiming the lives of 22% of children aged 1 to 5 years old in 2019 (**2**). Both viral and bacterial pathogens contribute to this condition, which spreads through airborne droplets and bodily fluids. Manifesting with symptoms such as chest pain, headache, cough, and fever, pneumonia imposes a substantial burden on public health. In this analysis, I sought to address two questions. First and foremost, my objective was to explore whether infant diet, particularly the timing of weaning off the breast milk, influenced infants being hospitalized with pneumonia. Secondly, I sought to explore additional factors beyond infant diet that could influence the risk of infant pneumonia hospitalization. To achieve these objectives, I utilized data from The National Longitudinal Survey of Youth, collected by the U.S. Bureau of Labor Statistics. This dataset comprises information gathered through interviews conducted with 3,470 mothers between 1979 and 1986 (**1**). By analyzing this rich dataset, we aimed to gain insights into the determinants of infant pneumonia hospitalization and contribute to our understanding of factors affecting child health outcomes. The variables in the dataset and their descriptions can be found in **Table 1**.

During the initial data exploration phase, summary statistics were computed for all variables. On average, infants were around 9-10 months old when hospitalized with pneumonia, while mothers had a mean age of 21.6 years at childbirth (**Table 2**). Infants were also typically weaned off breastmilk at an average age of 1.93 months and introduced to solid food at an average age of 1.12 months. Notably, only 2% of infants in the dataset were hospitalized with pneumonia. This indicates a heavily censored dataset, as evident in the ever so slight decrease followed by a plateau observed in the Kaplan-Meier survival curve presented in **Figure 2**. Turning our attention to the cumulative hazard curve in **Figure 2**, we confirm that the risk of getting hospitalized with pneumonia increases over time. Additionally, boxplots were generated to explore relationships between variables. Although most did not reveal significant trends, a trend can be seen regarding maternal age. The interquartile range was lower for mothers with a positive hospitalization status, suggesting a potential association between younger maternal age and higher rates of pneumonia hospitalization (**Figure 1**). This insight will be considered in subsequent analyses.

Statistical Analysis:

Both non-parametric methods and a backward model selection was conducted using the Efren method to determine significant covariates in our model. The stratified tests outlined in **Table 3.2** indicate that race and region, both categorical variables with multiple levels, were deemed insignificant and subsequently removed from the model. Then, non-parametric tests, Wilcoxon and Log-rank tests, revealed that smoke, the month of solid food introduction (sfmonth), number of siblings, and mother's age were significant covariates. Subsequently, a backward model selection further confirmed these findings, with the exception that the month of weaning off breast milk replaced sfmonth as a significant variable (**Table 3.1**). Given the focus on infant breast milk diet, the final set of significant covariates included mother's age, smoking habits, number of siblings, and the month of weaning off breast milk. This decision was made to prioritize variables directly related to the variable of interest, with comfort in removing sfmonth as it was the third variable eliminated from the backward selection model (**Table 3.1**). Subsequently, we fitted the final model using the Exact method and derived coefficient estimates for our parameters to complete the final model equation, as presented in **Table 4**.

Our final model is as follows:

$$\mathbf{h(y|x) = ho(y)*exp(0.691*smoke + 0.390*nsibs - 0.123*mthage -0.211*wmonth)}$$

Prior to delving into the effects of each variable in our model, a comprehensive assessment of model adequacy was carried out. This involved conducting tests to verify the proportional hazard assumptions, assess the global hypothesis, and evaluate the goodness of fit. The proportional hazard assumption was checked for all variables in **Table 5**. We observed that all covariates have p-values greater than 0.05. As a result, we fail to reject the null hypothesis that there is no violation of proportionality, and we conclude that the covariates pass the proportional hazard assumption. Therefore, a time-dependent analysis is unnecessary, and we can confidently proceed with our Cox proportional hazard modelling.

Upon model fitting, the global null hypothesis test results presented in **Table 6** indicate significant chi-square values for the likelihood ratio, score, and Wald tests, all yielding p-values less than 0.0001. These findings support the usefulness of our model with the dataset. However, it's important to note that this test does not provide information on the goodness of fit. Therefore, to assess the goodness of fit, a likelihood ratio test was conducted. **Table 7** provides the p-value from the likelihood ratio test comparing the final model and the full model. Given a p-value greater than 0.05 was obtained, we fail to reject the null hypothesis, indicating that the additional variables in the full model do not provide a better explanation of the data compared to the reduced model. It suggests that there is no significant difference in fit between the final model and the full model, and the former works just as well as the latter. As a result, we will move forward with our final model.

Discussion:

The final model revealed compelling insights into the factors influencing infant pneumonia hospitalization. Firstly, mothers who smoked during pregnancy exhibited a relative risk of 1.995, indicating mothers who smoked during pregnancy increased infant's pneumonia hospitalization risk by a factor of 1.995. The 0.995 or 99.5% increase in risk for mothers who smoke during pregnancy underscores the detrimental impact of maternal smoking on infant health, highlighting the urgent need for smoking cessation interventions among pregnant women to mitigate this risk. Moreover, the observed 1.477 relative risk in each additional sibling increases an infant's pneumonia hospitalization risk by a factor of 1.477. The 0.477 or 47.7% increase per additional sibling suggests potential mechanisms related to maternal caregiving capacity and increased exposure to different external pathogens. Mothers with more children may face challenges in providing adequate care for each child, and larger family sizes could lead to greater transmission of infectious agents within the household environment.

Interestingly, the relationship between maternal age and infant pneumonia hospitalization risk was notable. With a relative risk of 0.885, for every year increase in age of the mother, indicates a 0.115 or 11.5% reduction in infant pneumonia hospitalization risk. This finding may reflect the benefits of maternal experience and maturity in child-rearing practices. Older mothers may possess greater knowledge and resources to safeguard their infants' health, thereby reducing the likelihood of pneumonia-related hospitalizations. Furthermore, the weaning off breastmilk month shows a relative risk of 0.810. This indicates that per one month the transition off breast milk is delayed, there is a 0.19 or 19% reduction in the infant pneumonia hospitalization risk. This finding underscores the importance of appropriate infant feeding practices in reducing the vulnerability to respiratory infections. Delaying the weaning off breast milk may allow infants' immune systems to mature contributing to overall respiratory health.

In addition to uncovering the effects of each covariate, the final model can serve as a predictive tool for various scenarios defined by different covariate values. For instance, consider an infant with a 22-year-old mother who smoked during pregnancy, has 10 siblings, and was weaned off breastfeeding at 6 months. Employing the final model allows us to predict the 5th percentile of survival time, representing the minimum survival time where there's a 95% survival rate. According to Figure 4, in the scenario described, the 5th percentile equates to a survival time of 1 month. This contrasts starkly with an infant with a 22-year-old mother who did not smoke during pregnancy, has no siblings, and was weaned off breastmilk at 9 months. In this latter case, the survival time remains consistently above 0.99 or 99% throughout the 12 months, indicating notably higher survivability. These results further show the importance of providing optimal conditions for infants for better health outcomes.

Conclusion:

The study's findings truly underscore the multifaceted nature of factors influencing infant pneumonia hospitalization. The timing of weaning off breast milk exhibited a noteworthy influence on infant pneumonia-related hospitalizations, wherein extending breastfeeding by one month was associated with a significant 19% reduction in risk. Additionally, factors including mother's age, number of siblings, and smoking habits during pregnancy also emerged as influential contributors to infant pneumonia-related hospitalizations. Understanding these factors is crucial for developing targeted interventions to reduce infant pneumonia incidence and improve overall child health outcomes. Conducting further research with more contemporary data would offer valuable insights. Given the evolving habits and cultural shifts over time, a replication of the study could validate our findings and possibly uncover additional factors influencing pneumonia incidence in infants.

Appendix:

Table 1: Variable Definitions

Variable	Description
chldage	Age child had pneumonia, months
hospital	Indicator for hospitalization for pneumonia (1=yes, 0=no)
mthage	Age of the mother, years
urban	Urban environment for mother (1=yes, 0=no)
alcohol	Alcohol use by mother during pregnancy (1=yes, 0=no)
smoke	Cigarette use by mother during pregnancy (1=yes, 0=no)
region	Region of the country (1=northeast, 2=north central, 3=south, 4=west)
poverty	Mother at poverty level (1=yes, 0=no)
bweight	Normal birthweight (>5.5 lbs.) (1=yes, 0=no)
race	Race of the mother (1=white, 2=black, 3=other)
education	Education of the mother, years of school
nsibs	Number of siblings of the child
wmonth	Month the child was weaned
sfmonth	Month the child on solid food
agepn	Duration child in the hospital for pneumonia, months

Table 2: Descriptive Statistics

Variables (n = 3470)	**Data Type**	**mean (sd) or n (%)**
chldage	Quantitative	9.84 (3.62)
hospital:	Categorical	
Yes		73 (2.10)
No		3397 (97.90)
mthage	Quantitative	21.64 (2.72)
urban:	Categorical	
Yes		2639 (76.05)
No		831 (23.95)
alcohol:	Categorical	
Yes		626 (18.04)
No		2229 (64.24)
N/A		615 (17.72)
smoke:	Categorical	
Yes		838 (24.15)
No		2285 (65.85)
N/A		347 (10.0)
region:	Categorical	
Northeast		517 (14.91)
North Central		868 (25.02)
South		1396 (40.24)
West		689 (19.83)
poverty:	Categorical	
Yes		3200 (92.19)
No		270 (7.81)
bweight	Categorical	
Yes		1248 (35.93)
No		2222 (64.07)
race	Categorical	
White		1933 (55.70)
Black		957 (27.57)
Other		580 (16.73)
education	Quantitative	11.44 (2.0)
nsibs	Quantitative	0.68 (86)
		1.93 (3.64)

wmonth	Quantitative	
sfmonth	Quantitative	1.12 (1.99)
agepn	Quantitative	7.86 (4.46)

Figure 1: Exploratory Boxplots

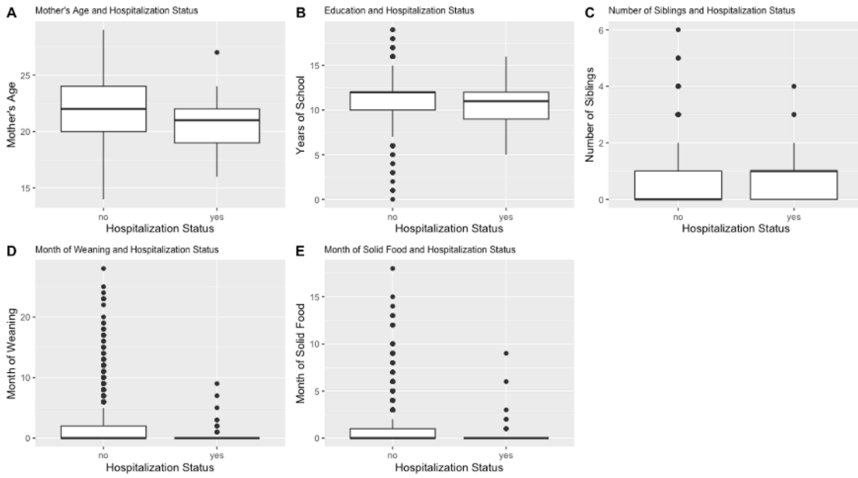


Figure 2: Kaplan-Meier Survival Curve

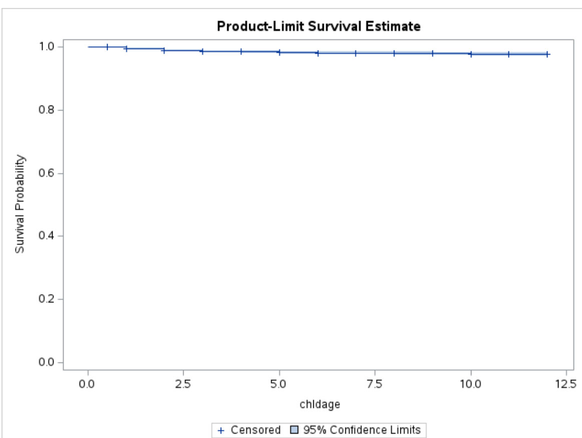


Figure 3: Cumulative Hazard Function

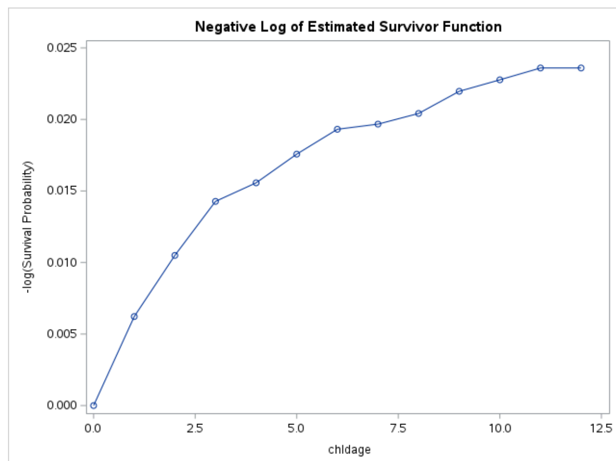


Table 3.1: Summary of Backward Elimination using Efren Test

Summary of Backward Elimination					
Step	**Effect Removed**	**DF**	**Number In**	**Wald Chi-Square**	**Pr > ChiSq**
1	poverty	1	12	0.0044	0.9472
2	agepn	1	11	0.0271	0.8694
3	sfmonth	1	10	0.1321	0.7163
4	alcohol	1	9	0.1518	0.6968
5	education	1	8	0.2922	0.5888
6	bweight	1	7	0.6678	0.4138
7	race	2	6	2.3960	0.3018
8	region	3	5	4.2764	0.2331
9	urban	1	4	1.7942	0.1804

Table 3.2: Non-Parametric Method for Significant Covariates Selection

Stratified Test of Equality over Group			
Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	0.9988	2	0.6069
Wilcoxon	0.7501	2	0.6873

*Effect of race within each region

Stratified Test of Equality over Group			
Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	6.0768	3	0.1079
Wilcoxon	4.6842	3	0.1964

*Effect of region within each race

Univariate Chi-Squares for the Wilcoxon Test				
Variable	Test Statistic	Standard Error	Chi-Square	Pr > Chi-Square
mthage	50.3273	22.6861	4.9214	0.0265
urban	6.6109	3.5988	3.3744	0.0662
alcohol	0.9074	4.0422	0.0504	0.8224
smoke	-13.9426	4.0110	12.0830	0.0005
poverty	1.3678	2.2543	0.3682	0.5440
bweight	-10.4896	4.0627	6.6664	0.0098
education	58.3296	16.8040	12.0491	0.0005
nsibs	-20.5241	7.1788	8.1738	0.0042
wmonth	103.8	31.3742	10.9434	0.0009
sfmonth	55.9783	16.9794	10.8691	0.0010
agepn	-16.7558	35.5695	0.2219	0.6376

Forward Stepwise Sequence of Chi-Squares for the Wilcoxon Test					
Variable	DF	Chi-Square	Pr > Chi-Square	Chi-Square Increment	Pr > Increment
smoke	1	12.0830	0.0005	12.0830	0.0005
sfmonth	2	20.9588	<.0001	8.8758	0.0029
nsibs	3	27.3159	<.0001	6.3571	0.0117
mthage	4	33.8755	<.0001	6.5596	0.0104
urban	5	35.6282	<.0001	1.7526	0.1855
wmonth	6	36.5693	<.0001	0.9411	0.3320
education	7	37.1228	<.0001	0.5535	0.4569
bweight	8	37.5015	<.0001	0.3787	0.5383
alcohol	9	37.5654	<.0001	0.0639	0.8005
agepn	10	37.5990	<.0001	0.0336	0.8546
poverty	11	37.6004	<.0001	0.00147	0.9694

Univariate Chi-Squares for the Log-Rank Test				
Variable	Test Statistic	Standard Error	Chi-Square	Pr > Chi-Square
mthage	50.7507	22.9609	4.8855	0.0271
urban	6.6517	3.6367	3.3455	0.0674
alcohol	0.8941	4.0877	0.0478	0.8269
smoke	-14.1056	4.0503	12.1289	0.0005
poverty	1.3835	2.2769	0.3692	0.5434
bweight	-10.6117	4.1048	6.6833	0.0097
education	58.7661	17.0035	11.9447	0.0005
nsibs	-20.7019	7.2499	8.1538	0.0043
wmonth	104.9	31.8598	10.8341	0.0010
sfmonth	56.5906	17.2342	10.7822	0.0010
agepn	-16.9028	35.9842	0.2206	0.6385

Forward Stepwise Sequence of Chi-Squares for the Log-Rank Test					
Variable	DF	Chi-Square	Pr > Chi-Square	Chi-Square Increment	Pr > Increment
smoke	1	12.1289	0.0005	12.1289	0.0005
sfmonth	2	20.9522	<.0001	8.8234	0.0030
nsibs	3	27.3295	<.0001	6.3773	0.0116
mthage	4	33.8976	<.0001	6.5680	0.0104
urban	5	35.6449	<.0001	1.7473	0.1862
wmonth	6	36.5744	<.0001	0.9296	0.3350
education	7	37.1127	<.0001	0.5382	0.4632
bweight	8	37.5098	<.0001	0.3971	0.5286
alcohol	9	37.5722	<.0001	0.0624	0.8028
agepn	10	37.6108	<.0001	0.0386	0.8442
poverty	11	37.6123	<.0001	0.00148	0.9693

*Tests with race and region removed

Table 4: Analysis of Maximum Likelihood Estimates using the Exact Method

Analysis of Maximum Likelihood Estimates						
Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
smoke	1	0.69088	0.23488	8.6517	0.0033	1.995
nsibs	1	0.39032	0.12150	10.3204	0.0013	1.477
mthage	1	-0.12271	0.04954	6.1348	0.0133	0.885
wmonth	1	-0.21101	0.08134	6.7304	0.0095	0.810

Table 5: Proportional Hazard Assumption Test

Supremum Test for Proportionals Hazards Assumption				
Variable	Maximum Absolute Value	Replications	Seed	Pr > MaxAbsVal
smoke	0.5295	1000	1614498530	0.6170
nsibs	0.3924	1000	1614498530	0.7910
mthage	0.3698	1000	1614498530	0.8970
wmonth	0.6936	1000	1614498530	0.3600

Table 6: Testing Global Hypothesis

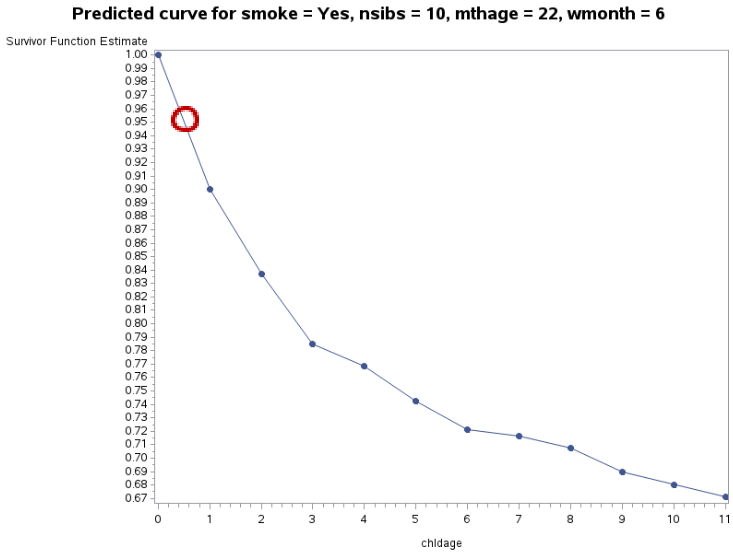
Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	39.3388	4	<.0001
Score	34.4374	4	<.0001
Wald	31.3294	4	<.0001

Table 7: Likelihood Ratio Test (Final model vs Full model): R output

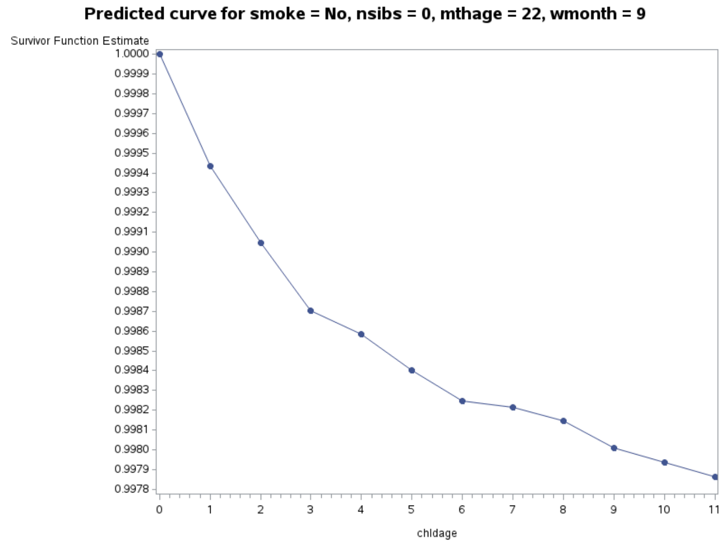
```
Likelihood ratio test

Model 1: Surv(chldage, hospital) ~ smoke + nsibs + mthage + wmonth
Model 2: Surv(chldage, hospital) ~ mthage + urban + alcohol + smoke +
    poverty + bweight + education + nsibs + wmonth + sfmonth +
    agepn + factor(race) + factor(region)
  #Df  LogLik Df Chisq Pr(>Chisq)
1   4 -568.69
2  16 -563.59 12 10.19     0.5993
```

Figure 4: Predictive Survival Rates Curve and Table



Obs	smoke	nsibs	mthage	wmonth	chldage	s	lcl	ucl	H	lH	uH
1	1	10	22	6	0	1.00000	.	.	0.00000	.	.
2	1	10	22	6	1	0.90009	0.69915	1	0.10526	0.009550	1.16026
3	1	10	22	6	2	0.83731	0.54828	1	0.17756	0.016358	1.92740
4	1	10	22	6	3	0.78530	0.44191	1	0.24169	0.022392	2.60866
5	1	10	22	6	4	0.76819	0.41033	1	0.26372	0.024461	2.84320
6	1	10	22	6	5	0.74250	0.36595	1	0.29773	0.027654	3.20544
7	1	10	22	6	6	0.72099	0.33147	1	0.32713	0.030412	3.51872
8	1	10	22	6	7	0.71663	0.32476	1	0.33320	0.030982	3.58344
9	1	10	22	6	8	0.70769	0.31133	1	0.34575	0.032160	3.71710
10	1	10	22	6	9	0.68964	0.28541	1	0.37159	0.034589	3.99201
11	1	10	22	6	10	0.68046	0.27282	1	0.38499	0.035847	4.13473
12	1	10	22	6	11	0.67116	0.26047	1	0.39875	0.037139	4.28133



Obs	smoke	nsibs	mthage	wmonth	chldage	s	lcl	ucl	H	lH	uH
1	0	0	22	9	0	1.00000	.	.	0	.	.
2	0	0	22	9	1	0.99943	0.99863	1	.000565169	.000136218	.002344889
3	0	0	22	9	2	0.99905	0.99772	1	.000953338	.000235819	.003854033
4	0	0	22	9	3	0.99870	0.99691	1	.001297624	.000324343	.005191500
5	0	0	22	9	4	0.99859	0.99663	1	.001415914	.000354775	.005650941
6	0	0	22	9	5	0.99840	0.99620	1	.001598518	.000401727	.006360698
7	0	0	22	9	6	0.99825	0.99583	1	.001756373	.000442275	.006974955
8	0	0	22	9	7	0.99821	0.99575	1	.001788984	.000450647	.007101929
9	0	0	22	9	8	0.99815	0.99560	1	.001856350	.000467913	.007364690
10	0	0	22	9	9	0.99801	0.99527	1	.001995100	.000503427	.007906659
11	0	0	22	9	10	0.99794	0.99510	1	.002067038	.000521843	.008187613
12	0	0	22	9	11	0.99786	0.99493	1	.002140934	.000540743	.008476489

References

1. *Cohorts | National Longitudinal Surveys*. www.nlsinfo.org/content/cohorts.

2. World Health Organization: WHO. *Pneumonia in Children*. 11 Nov. 2022, www.who.int/news-room/fact-sheets/detail/pneumonia.

SAS Code:

```
/* Generated Code (IMPORT) */
/* Source File: pneumon.csv */
/* Source Path: /home/u63559379/Survival Analysis */
/* Code generated on: 2/8/24, 10:16 PM */

%web_drop_table(WORK.IMPORT);


FILENAME REFFILE '/home/u63559379/Survival Analysis/pneumon.csv';

PROC IMPORT DATAFILE=REFFILE
      DBMS=CSV
      OUT=WORK.IMPORT;
      GETNAMES=YES;
RUN;

PROC CONTENTS DATA=WORK.IMPORT; RUN;


%web_open_table(WORK.IMPORT);

data pneumon;
    set work.import;
    if alcohol = 0 then alcohol = 0;
    else alcohol = 1;
    if smoke = 0 then smoke = 0;
    else smoke = 1;
run;

proc corr data=pneumon;
run;

/* K-M estimation */
title;
proc lifetest data=pneumon method=km plots=(survival(cl),ls,lls)
      graphics outsurv=a;
      time chldage*hospital(0);
      symbol1 v=none color=black line=1;
      symbol2 v=none color=black line=2;
run;

/* Non-parametric method to select significant covariates (stratified
by region)*/
proc lifetest data=pneumon method=km plots=(survival(cl),ls,lls)
```

```
      graphics;
      time chldage*hospital(0);
      strata region/ group=race;
      test mthage urban alcohol smoke poverty bweight education nsibs
wmonth sfmonth agepn race region;
run;

/* Non-parametric method to select significant covariates (stratified
by race)*/
proc lifetest data=pneumon method=km plots=(survival(cl),ls,lls)
      graphics;
      time chldage*hospital(0);
      strata race/ group=region;
      test mthage urban alcohol smoke poverty bweight education nsibs
wmonth sfmonth agepn race region;
run;

/* Non-parametric method to select significant covariates without race
and region*/
proc lifetest data=pneumon method=km plots=(survival(cl),ls,lls)
      graphics;
      time chldage*hospital(0);
      test mthage urban alcohol smoke poverty bweight education nsibs
wmonth sfmonth agepn;
run;

/* Conduct backward model selection ... */
proc phreg data=pneumon;
class race region;
      model chldage*hospital(0)= mthage urban alcohol smoke poverty
bweight education nsibs wmonth sfmonth agepn race region
           /ties=efron selection=backward;
run;

/*Global Test and likelihood ratio test for final model*/
proc phreg data=pneumon;
      model chldage*hospital(0)= smoke nsibs mthage wmonth
           /ties=efron;
run;

/*To get -2logL for likelihood ratio test for full model*/
proc phreg data=pneumon;
class race region;
      model chldage*hospital(0)= mthage urban alcohol smoke poverty
bweight education nsibs wmonth sfmonth agepn race region
           /ties=efron;
run;
```

```
/* Fit the final model and check its proportional hazard assumption */
proc phreg data=pneumon;
model chldage*hospital(0)= smoke nsibs mthage wmonth/ ties=exact;
assess ph/resample;
run;

/* Predicted survival function for hypothetically worst conditions */
data covals;
     input smoke nsibs mthage wmonth;
cards;
1 10 22 6
run;

proc phreg data=pneumon;
     model chldage*hospital(0)=smoke nsibs mthage wmonth /ties=exact;
     baseline out=pred covariates=covals survival=s lower=lcl
upper=ucl cumhaz=H
lowercumhaz=lH uppercumhaz=uH;
run;

proc print data=pred;
run;

/* print out & plot the predicted survival rates and find out the
95-percentile survival time */
/* the smallest time y such that S(y)< or = .95 */
proc gplot data=pred;
     title "Predicted curve for smoke = Yes, nsibs = 10, mthage = 22,
wmonth = 6";
     symbol1 value=dot i=join;
     plot s*chldage H*chldage;
run;

/* Predicted survival function for hypothetically better conditions */
data covals2;
     input smoke nsibs mthage wmonth;
cards;
0 0 22 9
run;

proc phreg data=pneumon;
     model chldage*hospital(0)=smoke nsibs mthage wmonth /ties=exact;
     baseline out=pred2 covariates=covals2 survival=s lower=lcl
upper=ucl cumhaz=H
lowercumhaz=lH uppercumhaz=uH;
run;

proc print data=pred2;
```

```
run;

/* print out & plot the predicted survival rates and find out the
95-percentile survival time */
/* the smallest time y such that S(y)< or = .95 */
proc gplot data=pred2;
      title "Predicted curve for smoke = No, nsibs = 0, mthage = 22,
wmonth = 9";
      symbol1 value=dot i=join;
      plot s*chldage H*chldage;
run;
```

R Code:

```r
library(ggplot2)
library(cowplot)

plot1 <- ggplot(pneumonia, aes(x = hospital, y = mthage)) +
  geom_boxplot() +
  labs(x = "Hospitalization Status", y = "Mother's Age") +
  ggtitle("Mother's Age and Hospitalization Status") +
  theme(plot.title = element_text(size = 9))

plot2 <- ggplot(pneumonia, aes(x = hospital, y = education)) +
  geom_boxplot() +
  labs(x = "Hospitalization Status", y = "Years of School") +
  ggtitle("Education and Hospitalization Status") +
  theme(plot.title = element_text(size = 9))

plot3 <- ggplot(pneumonia, aes(x = hospital, y = nsibs)) +
  geom_boxplot() +
  labs(x = "Hospitalization Status", y = "Number of Siblings") +
  ggtitle("Number of Siblings and Hospitalization Status") +
  theme(plot.title = element_text(size = 9))

plot4 <- ggplot(pneumonia, aes(x = hospital, y = wmonth)) +
  geom_boxplot() +
  labs(x = "Hospitalization Status", y = "Month of Weaning") +
  ggtitle("Month of Weaning and Hospitalization Status") +
  theme(plot.title = element_text(size = 9))

plot5 <- ggplot(pneumonia, aes(x = hospital, y = sfmonth)) +
  geom_boxplot() +
  labs(x = "Hospitalization Status", y = "Month of Solid Food") +
  ggtitle("Month of Solid Food and Hospitalization Status") +
  theme(plot.title = element_text(size = 9))

plot6 <- ggplot(pneumonia, aes(x = hospital, y = agepn)) +
  geom_boxplot() +
  labs(x = "Hospitalization Status", y = "Duration of Hospitalization") +
```

```r
  ggtitle("Duration of Hospitalization and Hospitalization Status") +
  theme(plot.title = element_text(size = 9))

plot_grid(plot1, plot2, plot3, plot4, plot5, labels = "AUTO")
```

```r
library(survival)
library(lmtest)

## Loading required package: zoo

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric

# Fit the final model (reduced model)
final_model <- coxph(Surv(chldage, hospital) ~ smoke + nsibs + mthage +
wmonth, data = pneumon)

# Fit the full model (final model)
full_model <- coxph(Surv(chldage, hospital) ~ mthage + urban + alcohol +
smoke + poverty + bweight + education + nsibs + wmonth + sfmonth + agepn +
factor(race) + factor(region), data = pneumon)

# Perform the likelihood ratio test
lr_test <- lrtest(final_model, full_model)
lr_test
```