01

RESEARCH QUESTION

- Does the mother’s age, alcohol use during pregnancy, whether the baby had a normal birthweight and race have an effect on the pneumonia occurrence rate in babies ?

02

OUR DATASET

Dataset

- Dataset chosen is from R-package KMsurv
- The data is titled **pneumon**
- The data frame has 3470 rows and 15 columns
- The factors relevant to our research question are hospital(**hospital**),age of the mother(**mthage**), alcohol use by the mother during pregnancy(**alcohol**), cigarette use by the mother during pregnancy(**smoke**), normal birthweight(**bweight**), race of the mother(**race)**, age child is in the hospital for pneumonia(**agepn**).

03

WHY DOES THIS INTEREST US?

Well, because…

- Pneumonia is a leading cause of mortality in infants and young children worldwide, making it an important area of study.
- Understanding the risk factors can help in developing prevention strategies to reduce the incidence and severity of pneumonia in this vulnerable population.
- This research is particularly important due to the potential long-term health implications of pneumonia in infants, such as respiratory problems, and the burden it places on families and healthcare systems.

04

SIGNIFICANCE OF INVESTIGATION

This research signifies…

1. Evidence-Based Decision Making:

The research provides valuable data for healthcare policymakers to make informed decisions regarding prevention, diagnosis, and treatment of childhood pneumonia.

2. Guiding Vaccine Development and Use:

By understanding the prevalence and distribution of specific factors causing pneumonia, the study can guide the development of effective vaccines, such as those for Streptococcus pneumoniae and Haemophilus influenzae type b (Hib).

3. Improved Resource Allocation:

Identifying high-risk populations can help prioritize the allocation of resources for intervention and support.

4. Long-Term Health Impact:

Reducing pneumonia incidence and severity can have long-term benefits for child health and development, as well as overall community health.

05

LITERATURE REVIEW

Article: “Pneumonia” by UNICEF

Key Challenges

High Incidence

The greatest incidence rates occur in South Asia and West and Central Africa.

Air Pollution

Around half of childhood pneumonia deaths are associated with air pollution, with indoor air pollution having a greater impact than outdoor pollution.

Slow Progress

Since 2000, childhood pneumonia deaths have decreased by 54%, while deaths due to diarrhea have declined by 63%.

Poverty Related Factors

Linked to factors such as undernutrition, lack of access to safe drinking water and sanitation, indoor and outdoor air pollution, and inadequate access to health care.

Lack of Early Care Seeking

Timely care-seeking for children with acute respiratory infection (ARI) symptoms can reduce mortality.

Article: “Pneumonia” by UNICEF

Key Interventions according to GAPPD*

Protect

1. Exclusive breastfeeding provides infants with protection from diseases and guarantees safe and nutritious food.

2. It also supports strong immune systems.

Prevent

1. Immunization are effective in preventing pneumonia.
2. Improved air quality also helps reduce cases.
3. In addition, preventing HIV and treating HIV infections helps maintain the immune system

Treat

1. Early recognition of danger signs such as fast or difficult breathing and cough is crucial.

2. Most serious Acute Respiratory Infection cases can be treated with antibiotics that are inexpensive.

* (Integrated Global Action Plan for the Prevention and Control of Pneumonia and Diarrhoea)

06

METHODOLOGY AND ANALYSIS

Kaplan-Meier Estimate

A Kaplan-Meier estimate model is given by:

$$\hat{S}(t) = \prod_{t_j < t} (1 - d_j / Y_j)$$

where $t_i < t$,

t_j : time at which event occurs
d_j : number of events at t_j
Y_j : number of individuals alive and at risk of event before t_j

Nelson-Aalen Estimate

A Nelson-Aalen estimator model is given by:

$$\hat{H}(t) = \sum_{t_j<t} (d_j/Y_j)$$

where $t_i \leq t,\ \hat{H}(t) = 0$

When $t_i > t,$ the variance is estimated by $\sigma_H^2 = \sum_{t_j<t} d_j/(Y_j)^2$

t_j : time at which event occurs

d_j : number of events at t_j

Y_j : number of individuals alive and at risk of event before t_j

Cox Proportional Hazards model

- A Cox proportional hazards model assumes that;

$$h(t \mid Z) = h_0(t)\exp(\beta^T Z)$$

where $h_0(t)$ is the baseline hazard function, $Z = (z_1, \ldots, z_p)^T$ is a $p \times 1$ vector of covariates and $\beta = (\beta_1, \ldots, \beta_P)^T$ is a $p \times 1$ vector of regression coefficient. $\exp(\beta^T Z)$ is called the Cox multiplier or proportionality constant or relative risk.

- The partial likelihood function is given as:

$$L(\beta) = \prod_{i=1}^{D} \frac{C_i *}{\sum_{j \in R} C_j}$$

where, $C_j = \exp[\sum_{k=1}^{p} \beta_k Z_{ik}]$.

Cox PH Assumption

- Use of time-dependent covariates to test the adequacy of the proportional hazards assumption.
- Relationship between the log hazard and each covariate is linear.
- Time dependent model

$$h(t \mid Z) = h_0(t)\exp(\beta^T Z + \gamma g(t))$$

- In addition, testing whether or not γ is significantly from zero allows us the opportunity to evaluate the proportional hazards assumption

$$log\frac{(h(t \mid Z_1))}{(h(t \mid Z_2))} = (Z_1 - Z_0)(\beta + \gamma g(t)).$$

Model Diagnostics

Cox-Snell residuals

- To check the overall fit of the model, we use Cox- Snell residual, which is given by;

$$r_j = \hat{H}_0(T_j)\exp(\beta^T Z_j)$$

where $Z_j = (z_1, \ldots, z_p)^T$ is a $p \times 1$ vector of fixed-time covariates and r_j are censored samples from the exponential distribution given that the assumed Cox model holds and the estimates of $\hat{H}_0(t)$ and β are close to its true values.

Martingale residuals

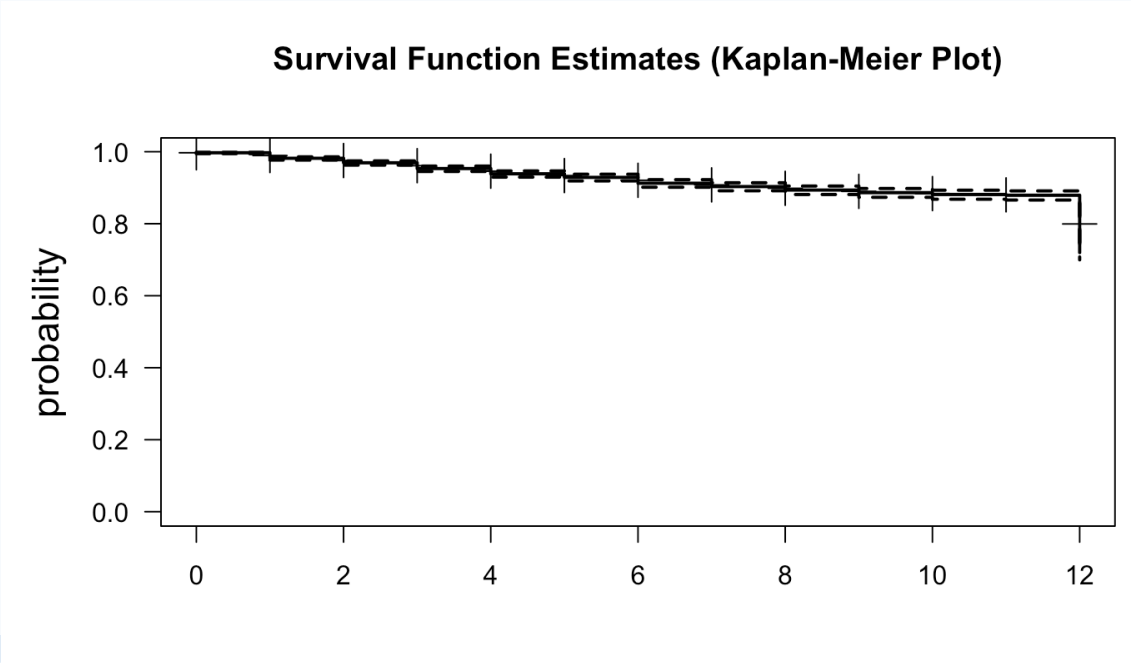
- Used to determine the function form of a covariates. If all covariates are time independent, with the usual RC (right-censoring), the martingale residual reduces to

$$\hat{M}_j = \delta_j - \hat{H}_0(T_j)\exp(\beta^T Z_j).$$

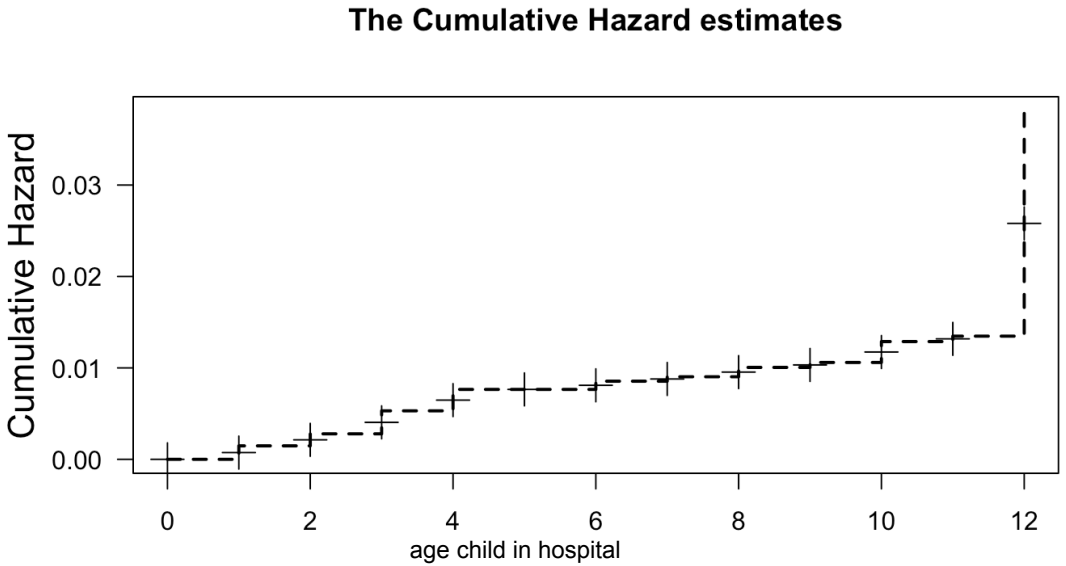
07

RESULTS AND CONCLUSIONS

Kaplan-Meier Estimation



Nalson-Aelon Estimation



Model Building (backward selection)

Variable	DF	Wald Chi Squared	P-Value	AIC
race	2	1.0016	0.29688	1097
mthage	1	0.2809	0.59802	1097
Alcohol	1	1.7956	0.18118	1099
Smoke	1	11.3569	0.00075	1101
bweight	1	4.41	0.03544	1108

Model Results

- Initial Model:
  Surv(agepn, hospital) ~ mthage + factor(alcohol) + factor(smoke) + factor(bweight) + factor(race)

- Final Model:
  Surv(agepn, hospital) ~ factor(smoke) + factor(bweight)

Variable	coef	exp(coef)	z	p-value
Smoke	0.8	2.2	3	0.001
bweight	0.5	1.7	2	0.03

- Likelihood ratio test=16 on 2 df, p=3e-04
  n= 3470, number of events= 73

Model Building (two-way interactions)

Variable	DF	AIC
bweight:race	2	1111
mthage:race	2	1111
smoke:race	2	1111
mthage:smoke	1	1112
alcohol:race	2	1112
mthage:alcohol	1	1113
alcohol:bweight	1	1113
smoke:bweight	1	1113
alcohol:smoke	1	1114
mthage:bweight	1	1117

Model Results

- Initial Model:
  Surv(agepn, hospital) ~ (mthage + factor(alcohol) + factor(smoke) + factor(bweight) + factor(race))^2

- Final Model:
  Surv(agepn, hospital) ~ mthage + factor(smoke) + factor(bweight) + mthage:factor(bweight)

Variable	coef	exp(coef)	z	p-value
mthage	-0.13	0.87	-2	0.048
smoke	0.76	2.14	3	0.001
bweight	-3.94	0.02	-2	0.042
mthage:bweight	0.21	1.24	2	0.021

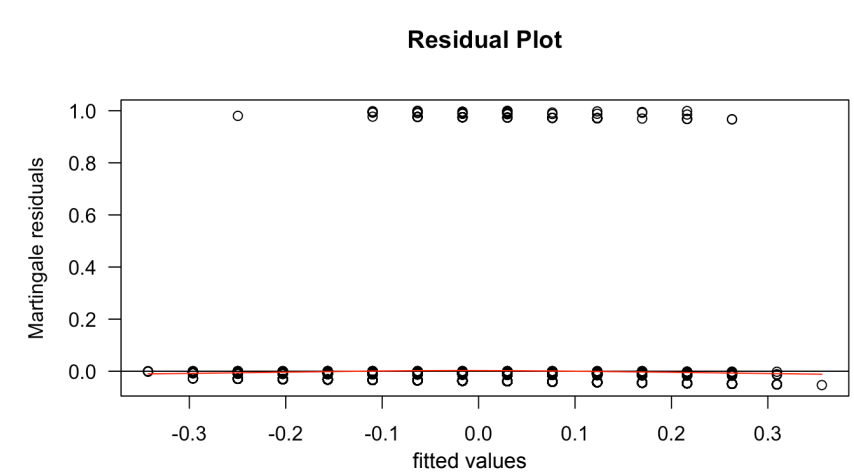
- Likelihood ratio test=22 on 4 df, p=2e-04
  n= 3470, number of events= 73

Assumption checking

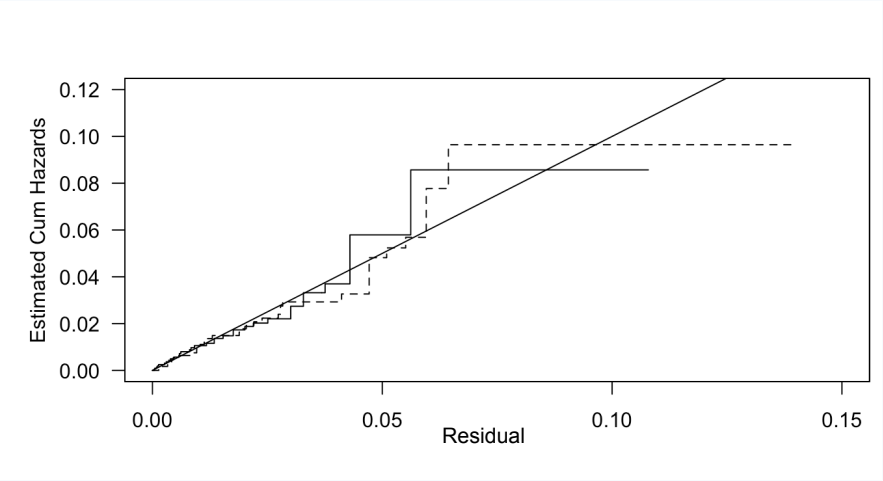
Variable	coef	exp(coef)	z	p-value
mthage	-0.14	0.87	-2	0.045
smoke	0.76	2.14	3.2	0.001
bweight	-4.33	0.01	-2.1	0.037
Z2t	0.16	1.17	0.5	0.602
mthage:bweight	0.22	1.24	2.3	0.019

- Z2t = bweight * ln(t)
- Likelihood ratio test=22 on 5 df, p=5e-04
  n= 25061, number of events= 73
- Assumption satisfied.

Martingale Residual plot



Cox-snell Residual plot



08

DISCUSSION

- The most significant factors were found to be the age of the mother at birth, cigarette use during pregnancy, whether the baby had a normal birth weight, and the interaction between the mother's age at birth and whether the baby had a normal birth weight.

- Those variables had the lowest p-values, and were used in the final model.

09

FUTURE WORK

- Research on pneumonia in babies is advancing on multiple fronts, aiming to improve prevention, diagnosis, and treatment.

- Research continues to focus on developing more effective vaccines against the most common pathogens causing pneumonia in infants.

- Early detection of pneumonia is essential for timely treatment and better outcomes. Researchers are exploring new diagnostic techniques, including advanced imaging technologies and biomarker detection, to diagnose pneumonia more quickly and accurately in infants.

- Other preventive measures include exclusive breastfeeding, improving indoor air quality, and reducing exposure to tobacco smoke.

REFERENCES

- https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8943783/
- https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3693495/
- https://www.ncbi.nlm.nih.gov/books/NBK536940/
- https://data.unicef.org/topic/child-health/pneumonia/