UHASSELT
KNOWLEDGE IN ACTION

Maastricht University

# Faculty of Sciences

## *School for Information Technology*

## Master of Statistics and Data Science

***Master's thesis***

***Innovative statistical analyses methods to quantify the performance of novel solar cooker appliances produced in resource-limited settings: A comparative approach***

**Baraka Rashidi**

Thesis presented in fulfillment of the requirements for the degree of Master of Statistics and Data Science, specialization Biostatistics

**SUPERVISOR :**

Prof. dr. Steven ABRAMS

2024
2025

Transnational University Limburg is a unique collaboration of two universities in two countries: the University of Hasselt and Maastricht University.

UHASSELT
KNOWLEDGE IN ACTION

**www.uhasselt.be**
Universiteit Hasselt
Campus Hasselt:
Martelarenlaan 42 | 3500 Hasselt
Campus Diepenbeek:
Agoralaan Gebouw D | 3590 Diepenbeek

# Faculty of Sciences
## *School for Information Technology*
Master of Statistics and Data Science

***Master's thesis***

***Innovative statistical analyses methods to quantify the performance of novel solar cooker appliances produced in resource-limited settings: A comparative approach***

**Baraka Rashidi**

Thesis presented in fulfillment of the requirements for the degree of Master of Statistics and Data Science, specialization Biostatistics

**SUPERVISOR :**

Prof. dr. Steven ABRAMS

## Acknowledgment

I would like to express my deepest gratitude to ALMIGHTY GOD for his unconditional love and for granting me the strength, wisdom, and miraculous moments throughout my studies and the completion of this thesis. He has played a significant role in my academic success.

I extend my sincere appreciation to my supervisors, Prof. dr. Steven ABRAMS, Prof. dr. Luc BIJNENS and Prof. dr. Didier KUMWIMBA, for their invaluable guidance, scholarly advice and critical suggestions throughout my master thesis, from the initial stages to the completion of my work. Their support has been instrumental in shaping this thesis to its final form.

I am delighted to the faculty and staff of UHasselt for creating an encouraging learning atmosphere and for providing the resources necessary for my academic endeavors. I would like to express my deepest appreciation to "VLIR-UOS" for awarding me the scholarship to study at UHasselt. I would also like to express my appreciation to the Sc4all team for their valuable collaboration towards the completion of this thesis.

My heartfelt gratitude goes to my family for their unwavering support, patience and love. Their encouragement has been a pillar of strength throughout this journey. I would like to acknowledge my classmates and colleagues who constantly supported me with their encouragement and advice. I would also like to acknowledge all individuals and institutions who contributed directly or indirectly to the successful completion of this thesis. Their contributions are sincerely appreciated.



Baraka Rashidi
June 17, 2025
Peperstraat 59,
Diepenbeek, Belgium.

# Contents

# Abstract

**Background:** The energy needed to cook accounts for 36% of the total energy usage worldwide. Approximately 2.4 billion people globally depend on non-renewable energy sources, such as solid fuels, for cooking, heating, and other domestic uses. Recently, the WHO reported that each year, 3.2 million people worldwide die from respiratory diseases caused by air pollution created by the use of solid fuels for cooking, especially in the Global South. An increase in the use of renewable energy sources, for cooking among other things, would therefore enhance sustainability in an effort to mitigate climate change.
**Objective:** This project aims to compare various survival analysis methods to evaluate the performance of various solar cooker appliances produced in resource-limited settings within the Sc4all project. Particularly, the study evaluates the time to reach threshold temperatures of 50 °C or 70 °C, where shorter times imply better performance of the cooker.
**Methods:** The study utilized survival analysis models to evaluate the effect of various covariates on the time to reach threshold temperatures of 50 °C or 70 °C, including Cox proportional hazards models and Accelerated Failure Time models. Model averaging was used to provide a comprehensive summary of the performance of different cooking appliances by accounting for uncertainty in the selection of the best modeling assumptions. The study focused solely on the time to reach threshold temperatures of 70 °C, as it is a plausible temperature to attain a boiling water temperature, according to the protocol.
**Results:** The findings from AFT models reveal that cooking performance in terms of the median time to reach a temperature threshold differs between different cookers and the use of a plastic bag to induce a greenhouse effect. Fixing the solar irradiance to $700W/m$ and the baseline water temperature to 20 °C, the prediction shows that Yama Dudo had superior performance, reaching 70 °C at shorter median time, followed by SK14. Among locally made cookers, Prototype 4 exhibited better performance, outperforming both other locally made cookers and a commercial Brother cooker.
**Conclusions:** Despite the clear differences between commercial devices and solar cooker prototypes developed in the context of the Sc4all project in terms of their performance, expressed conveniently in terms of the median time to reach a specific temperature threshold, the developement and implementation of solar cooking requires additional efforts in studying safety thereof. This study provides a first step towards a statistical modeling framework including refined survival analysis techniques for the analysis of solar cooker performance evaluation data with a more intuitive interpretation as compared to the established PEP protocol that is currently in place.

**Keywords :** *Solar energy, Censoring, Kaplan-Meier estimator, Cox PH model, AFT models, Model averaging.*

# 1 Introduction

## 1.1 Background

Energy is the ability to do work or cause changes in our daily relationships with our surroundings, and is also called a thermodynamic quantity. Due to the current challenges of the world and the need for clean energy generation, renewable energy sources have become dominant. Strong encouragement of research in renewable energy sources has been observed, especially after the oil crisis of 1973, with high fuel prices in a short period (Cuce and Cuce, 2013). Globally, energy demand is increasing due to the increased population and advances in technology (Kannan and Vakeesan, 2016).

The use of energy consumption shows an annual increase of 1% and 5% on average in developed and developing countries, respectively. This shows an increase in energy demand, which has led to the utilization of renewable energy sources at the global level over the last years. Renewable sources of energy occupy the throne as they contribute 14% to the world's energy demand, and this is expected to increase in the future (Herez et al., 2018).

In sub-Saharan African countries, charcoal is a significant source of energy and income, with more than 80% of urban households using it (Rose et al., 2022). Several reasons make charcoal a significant contributor to household income, especially in poor urban areas. It has the potential to cook and heat because it has a high heating content, less bulky, easier to transport, more accessible, and burns with less smoke compared to firewood (Zulu and Richardson, 2013). Cooking is essential for humans to prepare food for survival; according to Gorjian et al. (2022), the energy required for cooking accounts for 36% of the total energy consumption worldwide. Most countries still depend on carbon-based resources, including charcoal, firewood, and fossil fuels.

The use of charcoal in households has resulted in deforestation, air pollution, wood scarcity, natural disasters and energy shortages. Incomplete combustion of wood and other wood products at low amounts of oxygen and high temperatures produces a mixture of gases, liquid and solid particles known as wood smoke (Bede-Ojimadu and Orisakwe, 2020). The health impacts associated with the production process and the usage of charcoal are due to the smoke produced during incomplete combustion, which exposes humans to pollutants that lead to health risks and diseases such as cancer, heart disease, and lung disease (Idowu et al., 2023). According to a recent WHO report, more than 3.2 million people worldwide die from respiratory diseases each year due to air pollution, mainly indoors, due to the use of solid fuel for cooking, as shown in Figure 1.

Figure 1: *Visual representation of the complex challenges linked to charcoal use, encompassing health risks from outdoor air pollution, ecological harm driven by deforestation, and economic inefficiencies resulting from traditional production methods.*

Solar energy has become more significant among other renewable sources due to its reliability, cost-effectiveness and sustainability. It is a feasible and realistic solution in households for sustainable development in terms of health, environmental and economic challenges (Gorjian et al., 2022). Solar energy is estimated to fall on the Earth's surface with an average of 120 Petawatt, equivalent to the energy demand required in 20 years. International agencies predicted that in 2050, solar energy can supply up to 45% of the energy demand for the world (Herez et al., 2018). Therefore, it is important to pursue solar industry, which shows steady development and has become the best option for future energy demands because of its reliability and the fact that it is not limited to other renewable energy sources. Various research and innovations have been carried out and special devices have been developed to utilize solar energy as a power source for many industrial applications and technological advancements (Şen, 2004). A solar oven or cooker is a device that uses solar energy to do various tasks such as cooking food and performing other operations such as cleansing, sterilization and pasteurization.

This innovation emerged during the 18th century when the first experiment on solar cookers was done by the German Physicist, Tschirn-Hausen. In 1767, a French-Swiss physicist, Horace de Saussure, built a solar box to cook using solar energy where it reached a temperature of 88 °C (Herez et al., 2018). Afterward, many experiments were done to develop a strategy plan for a modern solar cooker. Currently, solar cookers are widely used in different styles but are mainly classified into solar panel cookers, solar parabolic and solar boxes (Cuce and Cuce, 2013). It has become an alternative as it encourages the adoption of sustainable practices and reduces dependence on non-renewable energy sources. Therefore, various numerical, analytical and experimental studies have been performed to enhance the power capacity of solar cookers (Herez et al., 2018).

The design of a solar cooker depends on the material used for construction, the climate of the area and the income level in both developing and developed countries. The household-level solar cookers are classified mainly into three classes, which are solar panel cookers, solar box cookers and solar parabolic cookers. The solar panel cooker is used mainly because of its low cost of manufacturing, as it uses cheap raw materials, unlike the solar box cooker. The solar box cooker is covered with a transparent glass cover so that the reflected sunlight is directed into the insulated box. It is one of the suitable cookers designed in the 1990s and early 2000s, consisting of a glass lid and the box is insulated to reduce heat loss from the box to the environment. Solar parabolic cookers can be self-constructed by using parabolas and large quantities of mirrors, but are not mainly used due to different obstacles, such as the availability of the raw materials used, affordability and safety usage. It is one of the best because it is heated to a maximum temperature within a short time, it contains a stand for a cooking system supporting a pot inside, located at the center where heating is facilitated by the parabolic reflector.

The Solar Cookers for All (Sc4all) project, which is financed by the Flemish government through a VLIR-UOS SI project (project number: CD2023SIN371A104) entails an interuniversity collaboration between the UHasselt and the University of Lubumbashi in the Democratic Republic of Congo (DRC). The Sc4all project aims to develop different solar cooker appliances that are locally produced based on available materials and in compliance with the specific needs of the local communities. The results presented in this master's thesis work rely on data collected within this project, in particular, for performance evaluation experiments involving solar cookers.

## 1.2 Objective of the Study

The main objective of this study is to compare different survival models to evaluate the performance of various solar cooker appliances produced in resource-limited settings within the Sc4all project. The performance of the solar cooker in each experiment was evaluated based on the time taken to reach threshold temperatures of 50 °C or 70 °C, where shorter times imply better performance of the cooker.

## 1.3 Research Questions

I. Which prototype(s) are the most effective compared to commercial devices in terms of the time to achieve the temperature threshold?

II. What are the key similarities and differences of the different approaches considered in addressing the performance of the existing solar cooker appliances?

III. What insights can be derived from the analyzed methods to improve the design and implementation of future solar cooker experiments?

## 1.4 Significance of the Study

The previous master thesis study of Kauki (2024) relied on the use of the Cox proportional hazards (PH) model to evaluate the performance of various solar cooker appliances produced in resource-limited settings within the Sc4all project. This study will contribute to the primary goal of the ongoing Sc4all project of designing low-cost solar cookers by providing an in-depth comparison of the different survival analysis approaches in quantifying the performance of novel solar cookers. By doing so, a thorough evaluation of the advantages and disadvantages, including the underlying assumptions made, of different modeling methods will be performed.

The report is organized as follows: Section 2 describes the data and introduces the methodology employed in the study, along with the statistical software used. In Section 3, the results of the data analysis are presented and discussed based on the fitted models. Lastly, Sections 4, 5 and 6 includes summary of the findings, ethical considerations, societal relevance and stakeholder awareness, and recommendations related to the master thesis, respectively.

# 2 Methods and Materials

## 2.1 Data Description

This study utilized data from experiments conducted in Belgium by the Sc4all project at UHasselt. Data were collected for various solar cooking appliances on different test dates for three years (from July 2022 to September 2024), leading to a dataset of 1148 observations and 28 variables. The original experimental data consists of measurements at the beginning and end of 10-minute time intervals with regard to ambient temperature, temperature in the cooking pot, solar irradiance and wind speed, among other time fixed factors such as the use of a plastic bag (see Table 1 for a detailed description of the original variables). These measurements are subsequently converted into time-to-event data, with the primary endpoint being the time to reach a specific temperature threshold of 50 °C or 70 °C, essentially leading to interval-censored time-to-event data. As a result, a total of 221 observations were included in this study, with variables of interest presented in Table 1.

Table 1: *Description of the variables in the study. The table contains both categorical/dummy and continuous variables.*

| **Variable** | **Description** | **Details** |
|---|---|---|
| Cooker type | 12 solar cooker appliances used | 4 oven types, 7 parabolic types and 1 OnlyPot |
| Ta1 | Ambient temperature at the beginning | Degree Celsius (°C) |
| Ta2 | Ambient temperature at the end of experiment | Degree Celsius (°C) |
| T1 | Water temperature inside the pot at the beginning of the experiment | Degree Celsius (°C) |
| T2 | Water temperature inside the pot after the experiment | Degree Celsius (°C) |
| I1 | solar irradiance at the beginning | Watts/ square meter |
| I2 | solar irradiance at the end of experiment | Watts/ square meter |
| H1 | Time at the beginning | Minutes |
| H2 | Time at the end of the experiment | Minutes |
| Time | Time to reach a threshold temperature | Minutes |
| Event | Event (or censoring) indicator to reach a threshold temperature | 0 = No event, 1 = Event of interest |
| PlasticBag | usage of a plasticbag to capture the greenhouse effect | 0 = No plastic bag, 1 = Plastic bag present |

Consequently, several sets of instruments were used to measure the aforementioned variables: An electronic balance was used to measure the water load in a pre-wetted container; a digital thermometer was used to measure the water temperature inside the cooking pot;

a pyranometer to measure solar irradiance and an anemometer was used to measure the ambient temperature.

## 2.2 Solar Cooker Appliances

The study involves several cooker appliances, i.e., oven prototypes (1,2,5 and 6), parabolic prototypes (2,3 and 4), and commercially available solar cookers being the Brother, Fornelia, Yamo Dudo and SK14 (see Figure 3). Figure 2 displays a black cooking pot labelled as OnlyPot, which is exposed outdoors to receive direct sunshine, with the water within anticipated to reach temperatures of approximately 50°C to 70°C within 3 hours, although it encounters difficulties in maintaining high temperatures.

Figure 2: *Visual representation of a cooking black pot, referred to as OnlyPot, which was merely used for comparison purposes to the solar cookers.*

The performance of commercially available solar cookers was evaluated in the experiment for comparison purposes with the locally made solar cookers.

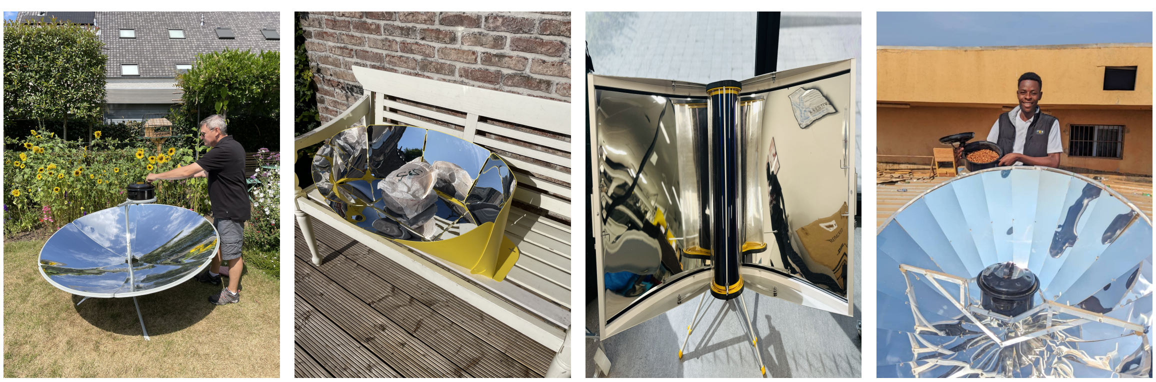

Figure 3: *Visual representation of the commercially available solar cookers included in the study, referred to as YamaDudo (left), Brother (middle), Fornelia (middle) and SK14 (right).*

Additionally, Figure 4 depicts the oven-type solar cookers designated as oven prototypes (1,2 and 5). Oven prototype 1 consists of a metallic frame box with a single reflective panel featuring aluminum foil. Oven prototype 2 is an improved version of "Oven prototype 1" which is enclosed in a cardboard box with four reflective panels featuring aluminum foil to optimize reflection from four solar angles. Oven prototype 5 consists of a wooden box featuring four reflective panels, which are covered with recycled soda cans and a cooking pot is supported by bricks, which also preserve heat.

Figure 4: *Visual representation of the locally made oven prototypes included in the study, referred to as oven prototype 1 (left), oven prototype 2 (middle) and oven prototype 5 (right).*

Furthermore, Figure 5 shows the locally made parabolic prototypes designated as prototypes 2, 3 and 4. Prototypes 2 and 3 comprise a steel frame featuring aluminum foil to reflect solar irradiation. These prototypes lack a supporting platform, complicating the adjustment towards the optimal solar direction. Prototype 4 was designed with a supporting stand that can be effortlessly adjusted towards the sun at right angles.

Figure 5: *Visual representation of the locally made parabolic prototypes included in the study, referred to as prototype 2/3 (left) and prototype 4 (right).*

## 2.3 Performance Evaluation Process

In November 2013, the American Society of Agricultural and Biological Engineers revised the protocol for evaluating and documenting solar cooker performance (ASAE S580.1). This standard aims to create uniformity and consistency in the terms and units used to define, test, rate, and assess solar cookers, their components and operating processes. It provides a standardized format for presentation and interpretation of test results, enhancing communication and providing a distinctive measure of thermal performance, allowing consumers to easily compare various designs when selecting a solar cooker. Moreover, this standard defines solar cooker as comprising the cooking vessel(s), supporting platform, heat transfer and retention surfaces, heat storage and transfer media, relevant pumps and controls, light transmitting and reflecting surfaces, as well as all necessary adjustments, supports and solar locating and tracking mechanisms that may be essential to a specific solar cooker (Funk, 2000). This standard ensures that the performance evaluation process (PEP) reduces the impact of all environmental factors that could affect the performance of the solar cooker, including wind, ambient temperature and solar irradiance (Mullick et al., 1987).

## 2.4 Survival Data Methods

Survival analysis is a domain in statistics studying time-to-event outcomes in the presence of censoring (Burzykowski, 2024) and (Klein and Moeschberger, 2003). Hence, survival analysis (also known as "time-to-event analysis") is used in several applied fields, including medicine, public health, social science and engineering (Qi, 2009). In essence, events are generally referred to as failures because they may be death, progression of disease, an incident, or others (Saikia and Barman, 2017). Survival data differs from other types of statistical data as the time until some specified event is not necessarily fully observed, leading to censoring, which can interrupt the observation of an event before it occurs (Lagakos, 1979). Due to the complexities provided by the censored data, various special statistical methods were developed to analyze survival data (Saikia and Barman, 2017). To address the main objective, the study utilized various survival analysis methods to examine the time taken by the solar cookers to reach threshold temperatures of 50 °C or 70 °C, accounting for other influencing variables.

### 2.4.1 Notations and Terminology

Let T represent a non-negative continuous random variable representing the time to a specific event of interest. Such time-to-event data is typically summarized using the survival and hazard function corresponding to the random variable T, for which estimators should accommodate censoring of observations (see Subsection 2.2.2 for more details, Qi, 2009). The survival function is defined as the probability that the survival time is larger than or equal to $t$, expressed as $S(t) = P(T > t)$ for $t \geq 0$. The probability density function (pdf) of $T$ is expressed as $f(t) = F'(t) = -S'(t)$ , where $F(t)$ is a cumulative distribution function. The hazard function is defined as the instantaneous risk of an event at time $t$ , which is given by

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t} = \frac{f(t)}{S(t)}, \quad t \geq 0. \tag{1}$$

The cumulative hazard function can be obtained from:

$$\Lambda(t) = \int_0^t \lambda(u)\, du = -ln[S(t)]. \tag{2}$$

#### 2.4.2 Censoring

Censoring is a fundamental feature in survival analysis, indicating the situation in which the event time is not fully observed for all subjects, although it is known to occur within a specific time interval (Ma, 2023). In essence, we denote $T \geq 0$ as the event time and $C \geq 0$ as the censoring time, assuming that $T$ is independent of $C$ (and non-informative censoring) (Lagakos, 1979). In a modern survival analysis, censoring can be broadly classified as left, right and interval censoring (Somasundaran, 2023). In particular, left censoring occurs when an observation exceeds its true value, that is $T < C$. In case of right censoring, the observed time is the censoring time, that is $T > C$. Interval censoring occurs when we only know that the true time of the event lies within a specific observed time interval that satisfies $C_L < T < C_U$ (Burzykowski, 2024). In most cases, interval-censored data are translated to right-censored data for convenience (Liu, 2012). Although in theory the data is interval-censored, we focus on right-censored observations obtained from experiments where the cooker never reached the threshold temperatures at any time throughout the experiment calendar period.

#### 2.4.3 Time-to-Event Data

The observations for the experiments are given by

$$(z_1, \delta_1, x_{1j}), (z_2, \delta_2, x_{2j}), \ldots, (z_n, \delta_n, x_{nj})$$

with

$$Z_i = \min(T_i, C_i) \quad \text{and} \quad \Delta_i = I(T_i \leq C_i), \tag{3}$$

where $T_i$ and $C_i$ are independent random variables for $i^{th}$ experiment for $i = 1, 2, \ldots, n$. Therefore, the observation $(Z_i, \Delta_i)$ is said to be uncensored if $\Delta_i = 1$, suggesting that the actual survival time has been observed. In contrast, an observation is said to be censored if $\Delta_i = 0$, suggesting that one has observed the censoring time $C_i$ and only has the partial information that $T_i > C_i$ (Gijbels, 2010). $x_{ij}$, for $j = 1, 2, \ldots, p$, is the covariate value of experiment $i$ corresponding to covariate $j$. It can either be a time-invariant covariate $x_{ij}$ or a time-varying covariate, denoted by $x_{ij}(t)$.

#### 2.4.4 Kaplan-Meier Estimate of the Survival Function

The Kaplan-Meier (KM) estimator is an estimator of the survival function $S(t)$ based on (right-)censored time-to-event data. It is also referred to as the product-limit estimator of the survival function and is a step function that jumps at uncensored event times while it remains constant whenever censored observations occur (Gijbels, 2010). More specifically, the KM estimator of $S(t)$ is defined as follows:

$$\hat{S}(t) = \prod_{j:t_{(j)} \leq t} \left(1 - \frac{d_j}{n_j}\right), \tag{4}$$

where $S(t) = S(t-1) \cdot P(\text{surviving up to time t})$, assuming $S(0) = 1$ for the observed times $t_1, t_2, \ldots, t_n$. Additionally, $t_{(1)}, t_{(2)}, \ldots, t_{(d)}$ are ordered event times for $d \leq n$, where $d_j$ is the number of observed temperatures of interest and because some events can be censored, $n_j$ represents the risk-set size at $t_{(j)}$, meaning the number of experiments that have not yet reached a threshold temperature before time $t_{(j)}$. As a result, $\frac{d_j}{n_j}$ is a non-parametric estimate of the hazard function, which represents the instantaneous risk of encountering a temperature of interest while accounting for tied events using the Efron method (Goel et al., 2010). In essence, to obtain a pointwise 95% confidence interval, the study essentially used a delta approach using Greenwood's formula to compute the variance of the KM estimator with $\log(-\log(S(t)))$ transformation. Furthermore, the log-rank test statistic was used to statistically compare two or more survival curves under the null hypothesis of no difference in survival probabilities between the different groups under consideration (Goel et al., 2010). This was chosen by expecting that the differences between cookers in their survival probabilities are equally weighted. The test statistic is given by

$$\text{Log-rank test statistic} = \sum_{g=1}^{G} \frac{(O_g - E_g)^2}{E_g}, \tag{5}$$

where $G$ equals the number of groups to compare, $O_g$ and $E_g$ are the total number of observed and expected events, respectively, and significance can be drawn by comparing the calculated value with the chi-squared of $G-1$ degrees of freedom under the null hypothesis (Goel et al., 2010).

### 2.5 Regression Approaches for Right-censored Time-to-Event Data

#### 2.5.1 Cox Proportional Hazard Model

The Cox PH model is defined on the scale of the hazard function (Burzykowski, 2024). It is a semi-parametric model with an unspecified baseline hazard function, and the covariates have a multiplicative effect on the hazard function (Saikia and Barman, 2017). The Cox

PH model is given by

$$\lambda(t|\mathbf{x_i}(t)) = \lambda_0(t)\exp\left(\sum_{j=1}^{p}\beta_j x_{ij}(t)\right), \tag{6}$$

where $\lambda(t|\mathbf{x_i}(t))$ is the hazard of an event of interest at time $t$ for the $i^{th}$ experiment (for $i = 1, 2, \ldots, n$) given the time-varying covariate $x_{ij}(t)$. $\lambda_0(t)$ is the baseline hazard for the subject and $\boldsymbol{\beta}$ is the $p \times 1$ vector of coefficients (Therneau et al., 2000). For covariate with two levels ($X = x_i$ and $X = x_j$) or a unit increase in a continuous covariate ($X = x$ and $X = x + 1$), the hazard ratio is constant over time and it is given by

$$\frac{\lambda(t|x_i)}{\lambda(t|x_j)} = \frac{\lambda_0(t)e^{x_i\beta}}{\lambda_0(t)e^{x_j\beta}} = \exp\left\{(x_i - x_j)\ \beta\right\} \equiv C. \tag{7}$$

The $\beta' s$ are estimated using the partial likelihood function proposed by Cox (1972). The partial likelihood is expressed as

$$\mathcal{L}(\beta) = \prod_{i=1}^{n}\prod_{t\geq 0}\left\{\frac{Y_i(t)\, r_i(\beta, t)}{\sum_j Y_j(t)\, r_j(\beta, t)}\right\}^{dN_i(t)}, \tag{8}$$

where $r_i(\beta, t)$ is the risk score for subject $i$, and $r_i(\beta, t) = \exp\left[x_i(t)\beta\right] \equiv r_i(t)$. In addition, the function $N_i(t)$ is the event counting process, which starts at 0 and stays 0 as long as subject $i$ has no event, and increases by 1 when an event occurs. The indicator function $Y_i(t)$ defines the risk-set, which is equal to 1 if the subject $i$ is under observation and at risk at time $t$, and 0 otherwise.

In particular, for time-invariant covarites ($x_{ij}(t) \equiv x_{ij}$), the PH model is given by

$$\lambda(t|\mathbf{x_i}) = \lambda_0(t)\exp\left(\sum_{j=1}^{p}\beta_j x_{ij}\right), \tag{9}$$

As a result, the final Cox PH model considered in this study is expressed as

$$\begin{aligned}\lambda(t \mid \mathbf{x_i}(t)) = \lambda_0(t)\exp\Big(&\sum_{j=1}^{6}\beta_j\,\text{Cooker}_{ji} + \beta_7\,\text{Baseline temperature}_i + \beta_8\,\text{Solar irradiance}_i(t)\\ &+ \beta_9\,\text{PlasticBag}_i + \beta_{10}\,(\text{PlasticBag}_i \cdot \text{Prototype } 5_i)\Big),\end{aligned} \tag{10}$$

where $\beta_1, \beta_2, \ldots, \beta_{10}$ are the coefficients corresponding to each covariate. *Cooker* represents dummy variables, where $Cooker_j = 0$ denotes the Prototype 4 cooker and $Cooker_j = 1$ indicates other remaining cookers included in the analysis. The *baseline temperature* is the temperature of the water at the beginning of a specific experiment. *PlasticBag* is a dummy variable for the use of a plastic bag covering the pot, with $PlasticBag = 1$ indicating the presence of a plastic bag and $PlasticBag = 0$ its absence. In terms of interaction, the study only considered (*PlasticBag vs. Prototype 5*) due to highly imbalanced plastic bag data in other cookers.

### 2.5.2 Accelerated Failure Time Models

The accelerated failure time (AFT) models have gained extensive attention in survival analysis and have become a crucial alternative to Cox models, as they are more explicit and direct in describing the covariate's effect on the event time than Cox models (Yang and Prentice, 2011). In contrast to the Cox PH model, the AFT model is a parametric distribution of the time to event instead of a semi-parametric one, and the effect of covariates acts multiplicatively on the logarithms of event times (Saikia and Barman, 2017). The general form of the AFT models is given by

$$\ln(T_i) = \mu + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \sigma \cdot \varepsilon_i, \tag{11}$$

where $\ln(T_i)$ is the natural log-transformed survival time for the $i^{th}$ subject; $x_1, \ldots, x_p$ are the time-invariant covariates with the coefficients $\beta_1, \ldots, \beta_p$; $\varepsilon_i$ is the *i.i.d* random error term that is assumed to follow a specific probability distribution; $\mu$ is the intercept that expresses the expected value of $\ln(T)$ when all covariates equals zero; and $\sigma$ is the scale parameter. Focusing on the right-censored data, the AFT models can be fitted using the maximum likelihood (ML) estimation approach. The likelihood function for right-censored time-to-event data, with the observed event times $t_1, \ldots, t_n$ is given by

$$L(\beta, \mu, \sigma) = \prod_{i=1}^{n} \{f_i(t_i)\}^{\delta_i} \{S_i(t_i)\}^{1-\delta_i} = \prod_{i=1}^{n} \left\{(\sigma \cdot t_i)^{-1} f_{\varepsilon_i}(w_i)\right\}^{\delta_i} S_{\varepsilon_i}(w_i)^{1-\delta_i}, \tag{12}$$

where $f_i(t_i)$ and $S_i(t_i)$ are the pdf and the survival function for the $i^{th}$ individual at time $t$, respectively.

Additionally, $f_{\varepsilon_i}(w_i)$ and $S_{\varepsilon_i}(w_i)$ are the density function and the survival function for the residuals, respectively, with

$$w_i = \frac{\ln(T_i) - \mu - \beta_1 x_{i1} - \cdots - \beta_p x_{ip}}{\sigma}. \tag{13}$$

The log-likelihood function is then;

$$l(\beta, \mu, \sigma) = \sum_{i=1}^{n} \left\{-\delta_i \ \ln(t_i) - \delta_i \ln(\sigma) + \delta_i \ln(f_{\varepsilon_i}(w_i)) + (1 - \delta_i)\ln(S_{\varepsilon_i}(w_i))\right\}. \tag{14}$$

ML estimates of the $(p+2)$ unknown parameters $\mu, \sigma$ and $\beta_1, \ldots, \beta_p$ are obtained by maximizing the loglikelihood function using, for example, the Newton-Raphson procedure (Saikia and Barman, 2017).

As a result, the survival function $S(t)$ in the AFT model can be expressed as

$$S(t \mid \mathbf{x}) = S_0\left(\frac{t}{\exp(\mathbf{x}\boldsymbol{\beta})}\right). \tag{15}$$

where $S_0(.)$ is the baseline survival function corresponding to the baseline covariate values. This reparameterization demonstrates that the survival time $t$ is rescaled by a quantity $\exp(\mathbf{x}\boldsymbol{\beta})$, referred to as the time acceleration or deceleration factor (or time ratio).

We considered the most frequently used AFT models, including Weibull, log-normal, log-logistic, generalized gamma and generalized F AFT models. The final AFT model employed, based on the residual assumptions of the model mentioned earlier, is mathematically expressed as

$$\begin{aligned}\ln(T_i) = \mu + &\sum_{j=1}^{6} \beta_j \,\text{Cooker}_{ji} + \beta_7 \,\text{Baseline temperature}_i + \beta_8 \,\text{Average solar irradiance}_i \\ &+ \beta_9 \,\text{PlasticBag}_i + \beta_{10} \,(\text{PlasticBag}_i \cdot \text{Prototype } 5_i) + \sigma \cdot \varepsilon_i,\end{aligned} \tag{16}$$

where $\ln(T_i)$ is the log-transformed of the time taken by a cooker to reach a threshold temperature. The coefficients $\beta_1, \ldots, \beta_{10}$ are the effects of the covariates, which are expressed as shortening ( $\beta_j < 0$) or lengthening ($\beta_j > 0$) the time to reach the threshold temperature. The interpretations will be based on the time ratios indicated in the Equation 15. Also, since solar irradiance is a time-varying covariate, we incorporated it in an average form for an experimental window.

**Generalized F model**

The generalized F is a family of AFT models, including Weibull, exponential, log-normal, log-logistic and generalized gamma models as special cases. The pdf of the generalized F with four parameters $\beta$, $\sigma$, $m_1$ and $m_2$ is obtained by

$$f_{GF}(t) = \frac{\exp(-\beta m_1/\sigma) \cdot t^{(m_1/\sigma)-1} \cdot (m_1/m_2)^{m_1}}{\sigma B(m_1, m_2)\left[1 + (m_1/m_2)\left(\exp(-\beta) \cdot t\right)^{1/\sigma}\right]^{(m_1+m_2)}}, \tag{17}$$

Where $q = (m_1^{-1} - m_2^{-1})(m_1^{-1} + m_2^{-1})^{-1/2}$ and $p = 2(m_1^{-1} + m_2^{-1})^{-1}$. Consequently, by fixing $q$ and $p$, generalized F becomes Weibull ($q = 1$, $p = 0$), log-normal ($q = 0$, $p = 0$), log-logistic ($q = 0$, $p = 1$) and generalized gamma ($q > 0$, $p = 0$).

**Generalized Gamma model**

The generalized gamma model is a subclass of the models included in the generalized F, which are Weibull, exponential and log-normal models. The survival function and the pdf of the generalized gamma distribution with three parameters $\mu$, $\kappa$ and $\sigma$ are given by

$$S(t) = \begin{cases} 1 - I(\gamma, u), & \text{if } \kappa > 0 \\ 1 - \Phi(w), & \text{if } \kappa = 0 \\ I(\gamma, u), & \text{if } \kappa < 0, \end{cases} \tag{18}$$

$$f(t) = \begin{cases} \frac{\gamma^\gamma}{\sigma t \sqrt{\gamma \Gamma(\gamma)}} \exp(w\sqrt{\gamma} - u), & \text{if } \kappa \neq 0 \\ \frac{1}{\sigma t \sqrt{2\pi}} \exp(-w^2/2), & \text{if } \kappa = 0, \end{cases} \tag{19}$$

where $\gamma = |\kappa|^{-2}$, $w = \text{sign}(\kappa) \cdot [\log(t) - \mu]/\sigma$, $u = \gamma \cdot \exp(|\kappa|w)$, $\Phi(w)$ is the standard normal cumulative distribution function and $I(\gamma, u)$ is the incomplete gamma function (Qi, 2009). As a result, by fixing $\kappa$ and $\sigma$, generalized gamma becomes Weibull ($\kappa = 1$), exponential ($\kappa = \sigma = 1$) and log-normal ($\kappa = 0$).

**Weibull model**

The Weibull model assumes that $\varepsilon_i$ follows a Gumbel distribution (*i.e.*, extreme value) with the pdf and survival function for the residuals $f_{\varepsilon_i}(w) = \exp(w - \exp(w))$ and $S_{\varepsilon_i}(w) = \exp(-\exp(w))$, respectively (Qi, 2009). Consequently, the event time $T$ is weibull distributed with parameters; $p = 1/\sigma$ and $\lambda = \exp(-\mu - \mathbf{X}'\boldsymbol{\beta})$ with a monotonic hazard function. In addition, an exponential model is a special case of the Weibull model with $\sigma = 1$. Moreover, the Weibull model is closed under both hazard-based and time-based transformations; thus, it can be expressed as a PH model with $\beta^* = -\beta/\sigma$.

**Log-normal model**

The log-normal model assumes that $\varepsilon_i$ follows a standard normal distribution with the pdf and survival function for the residuals $f_{\varepsilon_i}(w) = \frac{1}{\sqrt{2\pi}} e^{-w^2/2}$ and $S_{\varepsilon_i}(w) = 1 - \Phi(w)$ respectively (Saikia and Barman, 2017). Therefore, the event time $T$ is log-normally distributed with mean; $E(T) = \exp(\mu + \mathbf{X}'\boldsymbol{\beta}) \cdot \exp(\sigma^2/2)$ and variance; $\text{Var}(T) = \{E(T)\}^2 \cdot \{\exp(\sigma^2) - 1\}$.

**Log-logistic model**

The log-logistic model assumes that $\varepsilon_i$ follows a standard logistic distribution with the pdf and survival function for the residuals $f_{\varepsilon_i}(w) = \frac{p}{\lambda} \cdot \frac{(w/\lambda)^{p-1}}{[1+(t/\lambda)^p]^2}$ and $S_{\varepsilon_i}(w) = 1/(1 + e^w)$ respectively. As a result, the event time $T$ is log-logistically distributed with parameters; $p = 1/\sigma$ and $\lambda = \exp(\mu + \mathbf{X}'\boldsymbol{\beta})$ with a non-monotonic hazard function (Qi, 2009).

## 2.6 Choice of the Best Fitted Model

Model comparison strategies were employed to identify the most appropriate survival model for the data, encompassing both nested and non-nested models. A likelihood ratio test was used to compare two nested models. The test statistic is defined as:

$$-2\log L^{(1)} + 2\log L^{(2)} = -2\log\left(L^{(1)}/L^{(2)}\right), \tag{20}$$

where $L^{(1)}$ and $L^{(2)}$ are the maximized likelihoods for model 1 and model 2, respectively. Consequently, the test tests the null hypothesis that the additional parameters $q$ in model 2 are all zero, which follows an asymptotic chi-square distribution with $q$ degree of freedom. A nonsignificant result implies that the simpler model is adequate, whereas a significant result favors the complex model because of its added flexibility (Collett, 1994).

Furthermore, non-nested models including Weibull, log-normal, and log-logistic were compared using the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), defined respectively as:

$$\text{AIC} = -2\log(L) + 2k, \quad \text{BIC} = -2\log(L) + k\log(n), \tag{21}$$

where $L$ is the maximized likelihood, $k$ is the number of parameters for a particular model, and $n$ is the sample size. The model with the lowest AIC or/and BIC values is preferred, as it represents the optimal balance between model fit and complexity (Collett, 1994).

## 2.7 Model Diagnostics

Graphical methods for model diagnostics based on the residuals assessment are considered in the context of the widely used Cox PH model and for Anderson-Gill's generalization of that model. These residuals and/or their transforms are useful for evaluating the functional form of a covariate, the proportional hazards assumption, the leverage of each subject on the estimates $\beta' s$, and the model's goodness of fit given the subject (Therneau et al., 1990). This study examined a diagnostic assessment for the Cox PH model using martingale residuals, which assess the appropriate functional form of continuous covariates to be incorporated into the model. In addition to the Cox PH model, scaled Schoenfeld residuals were used to evaluate the proportional hazards assumption. Furthermore, the goodness of fit of the Cox PH and AFT models was assessed using the deviance and Cox-snell residuals, respectively (Collett, 1994).

### 2.7.1 Martingale Residuals

Martingale residuals are based on the difference between the counting process and the integrated intensity function (Therneau et al., 1990). These residuals were employed to assess the functional form of the continuous covariates to be incorporated in the Cox PH model. The martingale residual for the $i^{th}$ subject at each time $t$ is defined as

$$M_i(t) = N_i(t) - \int_0^t Y_i(s) e^{\beta' Z_i(s)} d\Lambda_0(s), \quad (i = 1, \ldots, n), \tag{22}$$

where $N_i(t)$ is the counting process representing the number of events experienced by the subject $i$ up to time $t$, $Y_i(s)$ is the at-risk indicator (1 if the subject is at risk at time $s$, 0 otherwise), $Z_i(s)$ is the covariate vector for the subject $i$, $\beta$ is the regression coefficient vector and $\Lambda_0(s)$ is the cumulative baseline hazard function. Martingale residuals are particularly useful for identifying non-linearity in covariates, as plotting these residuals against covariate values can reveal patterns indicating inappropriate functional forms. These residuals are skewed and bounded above one, making them best suited for detecting misspecification of continuous covariates in the Cox PH model.

### 2.7.2 Scaled Schoenfeld Residuals

The scaled Schoenfeld residuals are useful in assessing proportional hazards assumption, which is a crucial assumption in the Cox PH model. The Scaled Schoenfeld residual at the $k^{th}$ event time is defined as

$$\hat{r}^*_{s_k} = \hat{r}_{s_k}\,\mathrm{I}(\hat{\beta}, \hat{t}_{(k)}) = \hat{r}_{s_k}\,\mathrm{var}^{-1}(\hat{\beta}, \hat{t}_{(k)}), \tag{23}$$

where $\hat{r}_{s_k} = \int_{t_{k-1}}^{t_k} \sum_i (X_i(s) - \bar{X}(\hat{\beta}, s))\, dN_i(s)$, and $\hat{\beta}$ is the estimated coefficient. The plots of scaled Schoenfeld residuals are effective for detecting non-proportionality of estimated hazards in the fitted model across the covariate space (Grambsch and Therneau, 1994). In principle, the Schoenfeld residuals are expected to be time-independent; thus, a smoothed plot that shows a nonrandom pattern over time and/or systematic deviations from a horizontal line indicates the violation of the PH assumption. A test for independence between scaled Schoenfeld residuals and time was used to assess the PH assumption. A significant p-value indicates a violation of the assumption.

### 2.7.3 Cox-Snell Residuals

Cox-Snell residuals are typically used to assess overall goodness of fit in AFT survival models. We evaluated the presumed relation of unit exponentially distributed residuals for a good model fit, as well as under a specific model violation. This is done graphically using standard Cox-Snell residual plots. The Cox-Snell residual for the $i^{th}$ subject, denoted $r_{\mathrm{CS},i}$, is defined as

$$r_{\mathrm{CS},i} = \Lambda_i(t_i) = -\ln S_i(t_i) = -\ln S_\varepsilon(\varepsilon_i), \tag{24}$$

As a result, the modified Cox-Snell residuals are given by $r^*_{\mathrm{CS},i} = r_{\mathrm{CS},i} + 1 - \delta_i$, where $\delta_i = 0$ for censored observations and $\delta_i = 1$ for uncensored observations. In essence, for a properly fitted model, a plot of the Cox-Snell residuals $r^*_{\mathrm{CS}}$ against its estimated cumulative hazard $-\ln[S(r^*_{\mathrm{CS}})]$ is expected to show a straight line with zero intercept and unit slope (Cox and Snell, 1968).

## 2.8 Model Averaging Approach

Model averaging (MA) is a widely used and effective method for addressing model uncertainty and enhancing predictive accuracy. MA emphasizes the pooling of estimates by assigning higher weights to the better models, rather than depending on the model selection for a single model based on a goodness-of-fit criterion (Zhu et al., 2023). Using a single model might result in unreasonably small standard errors and narrow confidence intervals, as it ignores variation from other competing models (Anderson and Burnham, 2002). Consequently, the MA approach can provide better estimates and more reliable confidence intervals (Namata et al., 2008). As the distribution assumptions for the log time to reach a threshold temperature may be questionable, we estimated the performance of the solar cooker using three AFT models, including Weibull, log-normal and log-logistic and averaged across these models using AIC weights. The average estimate for the covariate is obtained by

$$\bar{\beta} = \sum_{m \in M} w_m \cdot \hat{\beta}_m,$$

with weights

$$w_m = \frac{\exp\left(-0.5\Delta_m\right)}{\sum_{k=1}^{M} \exp\left(-0.5\Delta_k\right)}. \tag{25}$$

where $\bar{\beta}$ is a weighted average of the estimate from model 16 with weights expressing the relative importance of the fitted AFT models. These weights are based on $\Delta_m$, which is the difference between the AIC value of the model $m$ and the AIC value of the model with the lowest AIC value. Furthermore, to use MA for statistical inference, the unconditional variance of the model-averaged estimate $\bar{\beta}$ for the large sample approximation is given by:

$$\text{Var}(\bar{\beta}) = \sum_{m \in M} w_m \left[\text{Var}(\hat{\beta}_m) + (\hat{\beta}_m - \bar{\beta})^2\right]. \tag{26}$$

This variance estimator accounts for within and between model variability in estimates, ensuring valid standard errors and confidence intervals that incorporate model uncertainty (Anderson and Burnham, 2002).

## 2.9 Software and Testing

All analyses were conducted in R Statistical Software version 4.4.3 (R Core Team, 2025) and SAS Studio for Academics (SAS Institute Inc., 2018). A significance level of 5% was used throughout the study.

# 3 Results

## 3.1 Exploratory Data Analysis

Table 2 shows the distribution of the events and censored events for each cooker involved in the study for the time to reach the threshold temperatures of 50 °C and 70 °C. Also, the table shows the distribution of experiments for whether or not the plastic bag was wrapped around the pot. It has been proposed that a minimum of 10 events should be observed for a covariate to be incorporated into multivariable models, such as the Cox PH model (Schober and Vetter, 2018). As a result, cookers with at least 10 events will be involved in further analysis, as shown by the blue colour. Also, observations for oven prototypes 1, 2 and 6 will be combined to make a single oven cooker designated as oven prototype 126. This approach gives more observations, which could make the estimates in the analysis more precise. Thus, the final model will only include the interaction term for the plastic bag with Prototype 5, as the distribution of plastic bags in other cookers is highly unbalanced. Furthermore, we will focus on the time required to reach a threshold temperature of 70 °C for the remainder of the analysis, as this is a reasonable endpoint to achieve a water boiling temperature according to the protocol.

Table 2: *The distribution of event occurrence and censoring for 50°C and 70°C for each cooker involved in the study, together with the distribution of the plastic bag (Yes or No).*

| | **50°C** | | **70°C** | | **Plastic Bag** | | |
|---|---|---|---|---|---|---|---|
| **Cooker** | **Censoring** | **Events** | **Censoring** | **Events** | **No** | **Yes** | **Total** |
| Yamo Dudo | 0 | 53 | 0 | 53 | 49 | 4 | 53 |
| Brother | 0 | 31 | 10 | 21 | 5 | 26 | 31 |
| OnlyPot | 1 | 5 | 6 | 0 | 3 | 3 | 6 |
| Oven prototype 1 | 0 | 1 | 0 | 1 | 0 | 1 | 1 |
| Oven prototype 2 | 0 | 5 | 1 | 4 | 0 | 5 | 5 |
| Prototype 2 | 0 | 10 | 5 | 5 | 1 | 9 | 10 |
| Prototype 3 | 1 | 3 | 2 | 2 | 0 | 4 | 4 |
| Prototype 4 | 0 | 18 | 0 | 18 | 1 | 17 | 18 |
| Prototype 5 | 0 | 14 | 2 | 12 | 9 | 6 | 14 |
| Prototype 6 | 0 | 8 | 2 | 6 | 8 | 0 | 8 |
| SK14 | 0 | 28 | 0 | 28 | 20 | 8 | 28 |
| Fornelia | 3 | 40 | 10 | 33 | 43 | 0 | 43 |

Figure 6 displays the KM curves of the survival function for the selected cookers individually compared to Prototype 4 without adjusting for other covariates information. The p-values from the log-rank test shown in the plots test for no significant difference in the time to reach a threshold temperature of 70 °C for the considered cookers at 5% level of significance. In particular, the plots demonstrate a highly significant difference between Yamo Dudo and Prototype 4 (p-value $< 0.0001$). Also, the plot shows that Brother differs significantly from Prototype 4 (p-value $= 0.0069$). Prototype 4 differs significantly from Prototype 5 (p-value $= 0.00024$). Oven prototype 126 differs significantly from Prototype 4 (p-value $< 0.0001$). Furthermore, SK14 and Fornelia are not significantly different from Prototype 4, with the p-values of 0.76 and 0.05, respectively. In addition to that, the KM curves for a threshold temperature of 50 °C are depicted in the Appendix (see Figure 11). Thus, because the KM method provides survival probability estimates without adjusting for other covariates, the study moved on to model-based approaches to incorporate the effects of those other covariates.

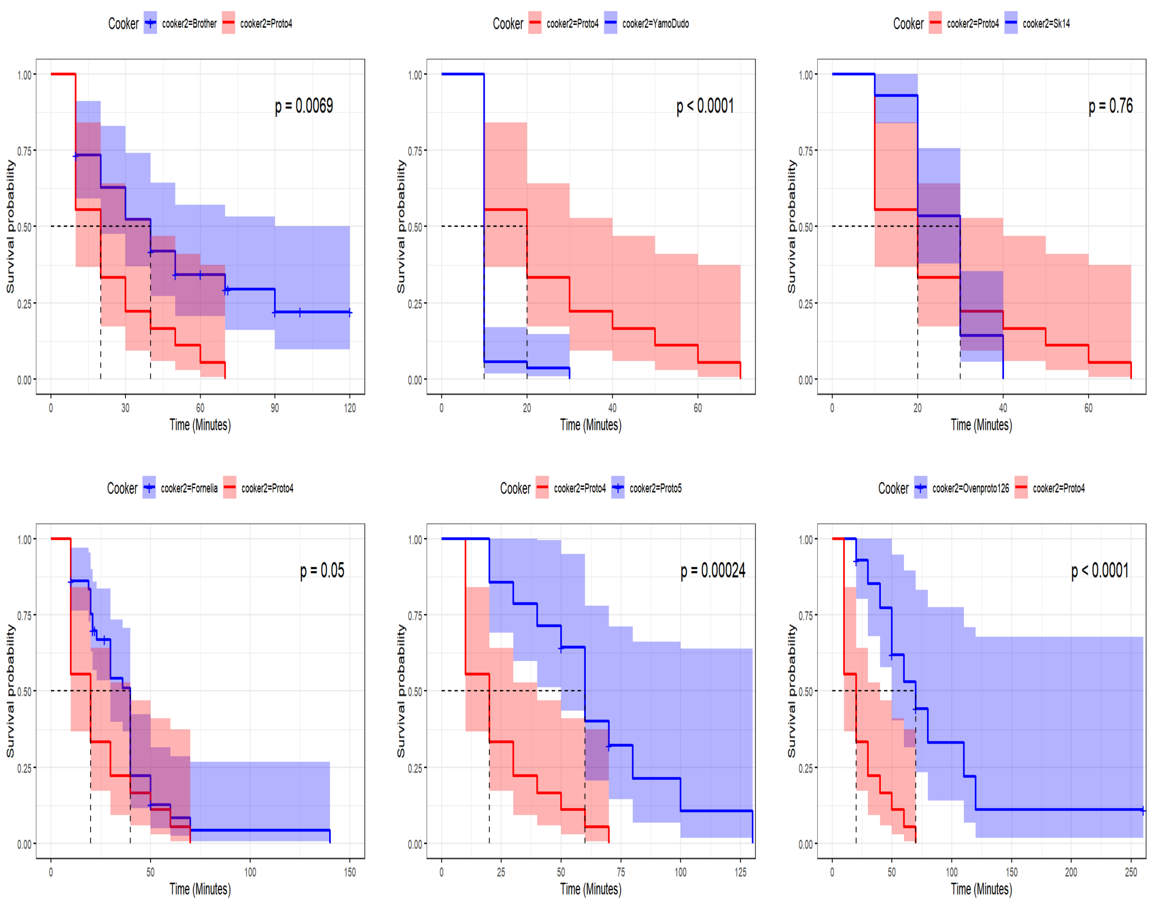


Figure 6: *Kaplan-Meier curves of the survival function with the 95% pointwise C.I for the cookers in the analysis. The two-sided p-values are based on a log-rank test comparing the survival functions of the cookers. The dashed lines show the medium time in minutes.*

## 3.2 Cox PH Model Diagnostics

### 3.2.1 Functional Form of Continuous Covariates

The martingale residuals from the null Cox PH model were plotted against each continuous covariate, such as Irradiation and the baseline temperature, to assess their appropriate functional form in the model. Figure 9 displays the smoothed plots for the time taken to reach a 70 °C. Thus, the smoothed plots indicate that the solar irradiance is reasonably linear and will be incorporated as a linear term in the model. In contrast, the baseline temperature shows that the smoothed lines are nonlinear.

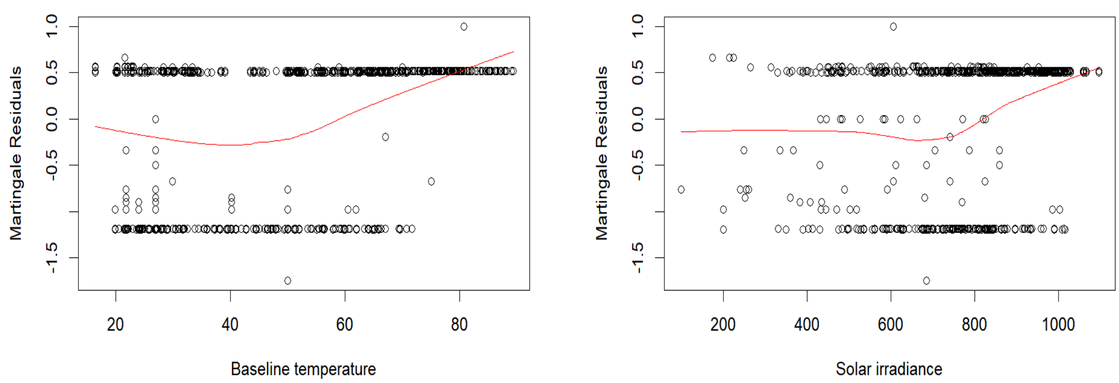


Figure 7: *The smoothed curves of the martingale residuals vs. continuous covariates included in the model. The red line is the smoothed line from a lowess function.*

Furthermore, Table 3 shows the significance test for nonlinearity using a Poisson regression approach. The result indicates that the solar irradiance can be included in a linear form, since the smoothing is not significant (p-value = 0.0664). Moreover, the baseline temperature is significantly different from linearity (p-value $< 0.0001$). Therefore, Figure 10 also indicates that the baseline temperature deviates from linearity and it has to be included as a nonlinear term in the model.

Table 3: *Summary of an approximate significance test of smoothed terms for non-linearity in the Poisson regression model. The (s) represents the smoothing parameter.*

| **Variable** | **Chisq** | **P-value** |
|---|---|---|
| s(Baseline temperature) | 29.18 | $< 0.0001$ |
| s(Solar irradiance) | 10.22 | 0.0664 |

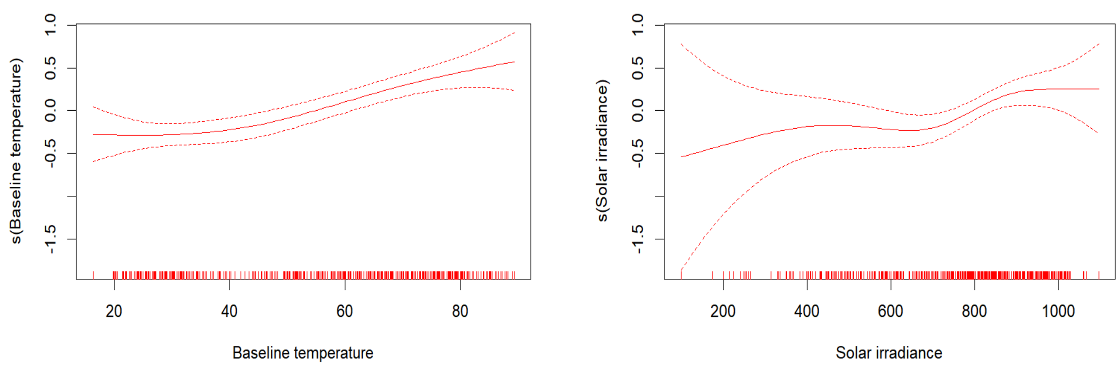


Figure 8: *The plots for assessing the functional form based on the Poisson regression approach for the continuous covariates.*

### 3.2.2 Proportional Hazards Assumption

Table 4 presents the summary of the Schoenfield test for proportional hazards assumptions to the covariates included in the Cox PH model. The results reveal that some covariates, such as Brother, Fornelia, Yamo Dudo, Oven prototype 126, baseline temperature and plasticBag, violate the assumption of proportional hazards due to significant p-values. This means that these covariate effects depend on time. In addition, the global test indicates an overall violation (p-value $< 0.0001$), which suggests that a Cox PH model may not adequately fit the provided data.

Table 4: *Summary of the Schoenfeld test results for proportional hazards assumptions. ns() represents the natural spline for baseline temperature with 4 degrees of freedom.*

| **Variable** | **Chisq** | **df** | **P-value** |
|---|---|---|---|
| Brother | 5.8660 | 1 | 0.0154 |
| Fornelia | 18.9220 | 1 | $< 0.0001$ |
| Oven prototype 126 | 4.6130 | 1 | 0.0317 |
| Prototype 5 | 0.4970 | 1 | 0.4810 |
| SK14 | 0.6910 | 1 | 0.4058 |
| Yamo Dudo | 11.4880 | 1 | 0.0007 |
| ns(Baseline temperature, df = 4) | 15.7150 | 4 | 0.0034 |
| Solar irradiance | 0.9020 | 1 | 0.3422 |
| PlasticBag | 13.0170 | 1 | 0.0003 |
| Prototype 5:PlasticBag | 1.1420 | 1 | 0.2853 |
| GLOBAL | 42.8100 | 13 | $< 0.0001$ |

Moreover, the scaled Schoenfeld residual plots for the covariates incorporated in the Cox PH model are presented in the Appendix (see Figure 12). The systematic departure from a horizontal line is indicative of the violation of the proportional hazards assumption. From the graphical inspection, some covariates indicate the violation of the proportional hazards assumptions as the lines deviate from the horizontal. For some cookers, the violation may be due to a few observations that tend to have wide intervals. As a result, these plots and the formal Schoenfeld test do not support the proportional hazards assumption. To address this, the study was expanded to incorporate accelerated failure time (AFT) models, providing a more suitable framework for time-to-event data when the proportional hazards assumption is not met. This allows for a more robust covariate's effects on the time to reach a threshold temperature.

### 3.3 Choice of an AFT model

Table 5 displays the AIC and BIC values for the fitted AFT models involved in model selection. The results indicate that log-logistic has the smallest AIC and BIC values, suggesting the best fit to the provided data relative to the Weibull and log-normal models. Since the generalised F and generalised gamma did not converge, we have opted for model selection based on AIC or/and BIC criteria. Therefore, it is shown that the log-logistic model provides the best fit.

Table 5: *The summary results of the AIC and BIC values for the fitted AFT models.*

| Model | logLik | df | AIC | BIC |
|---|---|---|---|---|
| Log-logistic | -639.48 | 11 | 1300.955 | 1337.181 |
| Log-normal | -644.18 | 11 | 1310.367 | 1346.593 |
| Weibull | -662.81 | 11 | 1347.622 | 1383.848 |

#### 3.3.1 Goodness of Fit for the Log-logistic model

The figures below displays the residual plots of the log-logistic model for the evaluation of the goodness of fit. As a result, Figure 9 implies that the KM of the data is likely to correspond to the survival probability of the residuals from the log-logistic model, as there is no substantial deviation of these functions. Thus, it suggests that the log-logistic model fits the data well. Moreover, Figure 10 shows the Cox-Snell residuals of the log-logistic model plotted against the cumulative hazards of the Cox-Snell residuals estimated using the KM approach. As a result, the observed Cox-Snell residuals behave like censored data from the unit exponential. Thus, the Cox-Snell residuals plot also indicates a good fit for the log-logistic model.

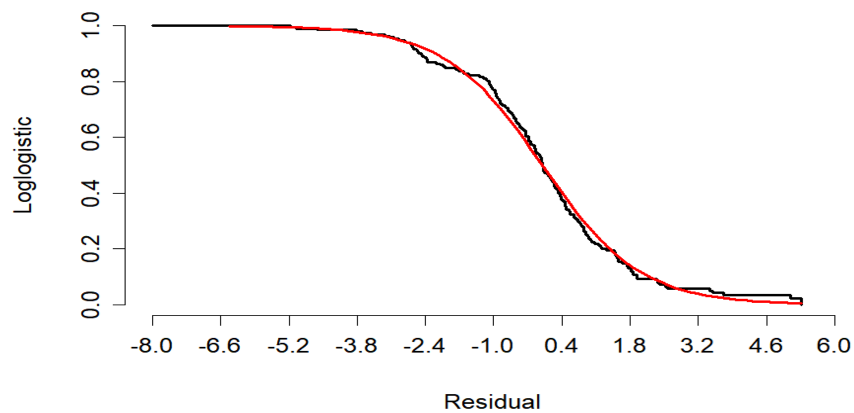


Figure 9: *Residual plot for assessing the goodness of fit for Log-logistic AFT model.*

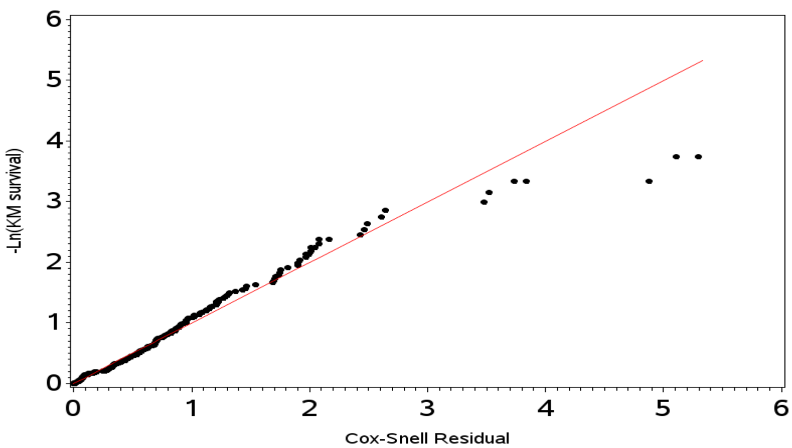


Figure 10: *Cox-Snell residual plot for assessing the goodness of fit for Log-logistic AFT model.*

### 3.3.2 Log-logistic Model Results

Table 6 summarises the parameter estimates for the fitted log-logistic model, specified in Equation 16. Since the interaction effect was not significant, we will focus our interpretation on the results obtained from the model without the interaction term. The results reveal that adjusting for other covariates in the model, the median time to reach the threshold of 70 °C is not significantly different for Fornelia as compared to the reference prototype being prototype 4 (p-value = 0.7504).

For Brother, the results show a significant difference from prototype 4 (p-value = 0.0009), and the estimated median time of 1.6409 indicates a longer time to reach 70 °C by 64% as compared to prototype 4, with a 95% confidence interval ranging from 1.2231 to 2.2011. Prototype 5 is significantly different from prototype 4 (p-value = 0.0026), and the estimated median time of 1.7071 indicates a longer time to reach 70 °C by 71% as compared to prototype 4, with a 95% confidence interval of 1.21 to 2.42. The oven prototype 126 has an estimated median time of 2.2146, which indicates a longer time to reach 70 °C by 121% as compared to prototype 4. This effect is highly significant (p-value $< 0.0001$) with a 95% confidence interval of 1.5607 to 3.1426. In contrast, the results show that Yamo Dudo and SK14 have shorter times compared to prototype 4. In particular, SK14 takes a median time of 0.6702, which indicates a shorter time to reach 70 °C by 33%. This effect is significantly different (p-value = 0.0131) with a 95% confidence interval of 0.4889 to 0.9194. Yamo Dudo takes a median time of 0.3910, which indicates a shorter time to reach 70 °C by 61%. This effect was highly significant (p-value $< 0.0001$) with a 95% confidence interval of 0.2937 to 0.5198.

Furthermore, one-degree Celsius increase in the baseline temperature leads to a decrease in median time to reach a temperature of 70 °C by 2%. This effect is highly significant (p-value $< 0.0001$) with a 95% confidence interval of 0.9715 to 0.9814. Also, $100W/m^2$ increase in the average solar irradiance leads to a decrease in median time to reach a temperature of 70 °C by 63%. This effect is highly significant (p-value $< 0.0001$) with a 95% confidence interval of 0.998 to 0.999. In addition, the presence of a plastic bag around the cooking pot decreases the median time by 19% as compared to when a plastic bag is absent. This effect is statistically significant (p-value = 0.0324) with a 95% confidence interval of 0.6678 to 0.9824.

Table 6: *Summary results of the fitted Log-logistic AFT model with the 95% confidence intervals (CI) for the time ratios (TR).*

| **Variable** | **Parameter** | **Estimate** | **TR** | **SE** | **P-value** | **95% CI** |
|---|---|---|---|---|---|---|
| Intercept | $\mu$ | 5.1855 | 178.675 | 0.3027 | $< 0.0001$ | [98.28, 322.05] |
| Brother | $\beta_1$ | 0.4951 | 1.6409 | 0.1499 | 0.0009 | [1.2231, 2.2011] |
| Fornelia | $\beta_2$ | -0.0571 | 0.9444 | 0.1794 | 0.7504 | [0.6647, 1.3422] |
| Oven prototype 126 | $\beta_3$ | 0.7952 | 2.2146 | 0.1785 | $< 0.0001$ | [1.5607, 3.1426] |
| Prototype 5 | $\beta_4$ | 0.5346 | 1.7071 | 0.1772 | 0.0026 | [1.2060, 2.4150] |
| SK14 | $\beta_5$ | -0.4002 | 0.6702 | 0.1614 | 0.0131 | [0.4889, 0.9194] |
| Yamo Dudo | $\beta_6$ | -0.9392 | 0.3910 | 0.1454 | $< 0.0001$ | [0.2937, 0.5198] |
| Baseline temperature | $\beta_7$ | -0.0239 | 0.9764 | 0.0026 | $< 0.0001$ | [0.9715, 0.9814] |
| Average solar irradiance | $\beta_8$ | -0.0013 | 0.9987 | 0.0003 | $< 0.0001$ | [0.9982, 0.9992] |
| PlasticBag (Yes) | $\beta_9$ | -0.2109 | 0.8096 | 0.0985 | 0.0324 | [0.6678, 0.9824] |

Furthermore, Table 7 presents the summary of the pairwise comparisons between cookers from the log-logistic model. The results reveal that there is a highly significant difference among the commercial solar cookers in the time it takes to reach a temperature of 70 °C (p-values $< 0.0001$). The results indicate that SK14 takes 71% longer median time as compared to Yamo Dudo, with 95% CI ranging from 1.2309 to 2.3877. Brother takes 145% and 420% longer median time than SK14 and Yamo Dudo, respectively.

Table 7: *Summary of the pairwise comparisons of the cookers on the median time to reach 70 °C.*

| **Contrast** | **TR** | **SE** | **P-value** | **95% CI** |
|---|---|---|---|---|
| Brother - Oven prototype 126 | 0.7406 | 0.164 | 0.2787 | [0.4465, 1.2294] |
| Brother - Prototype 5 | 0.9613 | 0.159 | 1.0000 | [0.5888, 1.5695] |
| Brother - SK14 | 2.4485 | 0.139 | $< .0001$ | [1.5946, 3.7570] |
| Brother - Yamo Dudo | 4.1970 | 0.135 | $< .0001$ | [2.7696, 6.3624] |
| Oven prototype 126 - Prototype 5 | 1.2975 | 0.171 | 0.3873 | [0.7667, 2.1974] |
| Oven prototype 126 - SK14 | 3.3035 | 0.150 | $< .0001$ | [2.0823, 5.2432] |
| Oven prototype 126 - Yamo Dudo | 5.6645 | 0.142 | $< .0001$ | [3.6628, 8.7742] |
| Prototype 5 - SK14 | 2.5463 | 0.136 | $< .0001$ | [1.6750, 3.8718] |
| Prototype 5 - Yamo Dudo | 4.3645 | 0.134 | $< .0001$ | [2.8856, 6.6093] |
| SK14 - Yamo Dudo | 1.7147 | 0.108 | $< .0001$ | [1.2309, 2.3877] |

### 3.4 Model Averaging Results

Table 8 presents a summary of the parameter estimates for the fitted AFT models as defined in Equation 16, together with the model-averaged estimates. The findings reveal that the estimates and standard errors in these fitted models do not differ substantially in most variables. The findings show that the model averaging approach produces estimates and standard errors that are very close to those obtained from the log-logistic model. This is expected because the model-averaged estimates are dominated by a log-logistic model with a weight of 0.99. Therefore, we considered the predictions of the median time to reach a temperature of 70 °C based on the log-logistic model as the best predictions for this study.

Table 8: *Summary of the parameter estimates and standard errors for different AFT models fitted to the provided data and the model-averaged estimates.*

| | **Log-logistic** | | **Log-normal** | | **Weibull** | | **MA** | |
|---|---|---|---|---|---|---|---|---|
| **Variable** | **Estimate** | **SE** | **Estimate** | **SE** | **Estimate** | **SE** | **Estimate** | **SE** |
| Intercept | 5.1855 | 0.3027 | 5.3509 | 0.3005 | 5.8128 | 0.3374 | 5.1870 | 0.3031 |
| Brother | 0.4951 | 0.1499 | 0.4539 | 0.1404 | 0.5526 | 0.1535 | 0.4948 | 0.1499 |
| Fornelia | -0.0571 | 0.1794 | -0.1100 | 0.1823 | -0.2287 | 0.1833 | -0.0575 | 0.1795 |
| Oven prototype 126 | 0.7952 | 0.1785 | 0.8203 | 0.1753 | 0.8634 | 0.1873 | 0.7954 | 0.1785 |
| Prototype 5 | 0.5346 | 0.1772 | 0.4762 | 0.1816 | 0.2621 | 0.1958 | 0.5341 | 0.1774 |
| SK14 | -0.4002 | 0.1614 | -0.4905 | 0.1667 | -0.7014 | 0.1699 | -0.4010 | 0.1616 |
| Yamo Dudo | -0.9392 | 0.1454 | -0.9586 | 0.1492 | -1.0919 | 0.1386 | -0.9394 | 0.1454 |
| Baseline temperature | -0.0239 | 0.0026 | -0.0244 | 0.0026 | -0.0251 | 0.0028 | -0.0239 | 0.0026 |
| Average solar irradiance | -0.0013 | 0.0003 | -0.0014 | 0.0003 | -0.0016 | 0.0003 | -0.0013 | 0.0003 |
| PlasticBag (Yes) | -0.2109 | 0.0985 | -0.2357 | 0.1015 | -0.1443 | 0.0950 | -0.2111 | 0.0986 |
| AIC | | 1300 | | 1310 | | 1347 | | — |
| WEIGHTS | | 0.991 | | 0.009 | | 0.000 | | — |

### 3.5 Prediction of the Median Survival Time

To determine the cooker with the best performance in the analysis, the predictions were based on the median time taken for a certain cooker to reach a temperature of 70 °C, where a shorter time implies better performance of the cooker. We performed the prediction by fixing the covariate values in the log-logistic model, in accordance with the ASAE S580.1 protocol. As a result, the average solar irradiance was assessed at $700W/m^2$, as the cooking power for each interval should be adjusted to a standard insolation of $700W/m^2$ (Funk, 2000). In addition, we considered the fixed baseline temperature of 20 °C and evaluated the median time to reach 70 °C from 20 °C. Furthermore, the prediction was based on whether or not the plastic bag was wrapped around the cooking pot, to gain insight into whether the usage of a plastic bag is potentially advantageous while cooking.

Table 9 provides a summary of the predicted median time to reach a temperature of 70 °C for the solar cookers, along with the prediction intervals (PI) based on asymptotic standard error estimates. It also shows how the time taken varies depending on whether or not the plastic bag is used. It is postulated that the use of a plastic bag reduces the time to reach 70 °C for all solar cookers compared to the absence of a plastic bag. The predictions indicate that Yama Dudo performs better compared to other commercially and locally made solar cookers, as it takes a median of 15 minutes to achieve a temperature of 70 °C when the plastic bag is used with the 95% PI ranging from 8 to 28 minutes and a median of 18 minutes without a plastic bag with the 95% PI ranging from 10 to 33 minutes. Moreover, for locally made cookers, Prototype 4 appears to perform better than other locally made cookers, as well as the Brother, which is a commercial solar cooker. As a result, Prototype

4 takes a median of 37 minutes to achieve a temperature of 70 °C when the plastic bag is used with the 95% PI ranging from 20 to 68 minutes and a median of 45 minutes without a plastic bag with the 95% PI ranging from 25 to 81 minutes.

Table 9: *Predicted median times to reach* 70 °C *(minutes) with 95% prediction intervals (PI) for different solar cookers under hypothetical conditions.*

| **Cooker** | **PlasticBag: Yes** | | **PlasticBag: No** | |
|---|---|---|---|---|
| | **Time (min)** | **95% PI** | **Time (min)** | **95% PI** |
| Yamo Dudo | 15 | [8, 28] | 18 | [10, 33] |
| SK14 | 25 | [12, 49] | 30 | [16, 59] |
| Fornelia | — | — | 43 | [22, 84] |
| Prototype 4 | 37 | [20, 68] | 45 | [25, 81] |
| Brother | 60 | [30, 118] | 74 | [38, 142] |
| Prototype 5 | 62 | [31, 126] | 77 | [39, 152] |
| Oven prototype 126 | 99 | [40, 164] | 80 | [50, 197] |

# 4 Discussion and Conclusion

The purpose of this study is to investigate different statistical analysis approaches to quantify the performance of several solar cooker appliances produced in resource-limited settings as part of the Sc4all project. The project focused on survival analysis techniques, which evaluate the time required for solar cookers to reach a specified threshold temperature, as specified in the protocol provided by ASABE. Survival analysis approaches offered a flexible and effective way to evaluate the performance of solar cookers, accommodating the complexity of the data collected in these performance experiments, and providing further insights compared to conventional analysis techniques that emphasize a single measure of performance through simple linear regression (Funk, 2000).

Consequently, several survival analysis approaches were employed to fulfill the objectives of this study. To account for potential factors that influence the performance of the solar cooker appliances such as solar irradiance and the baseline water temperature, this study focused on survival models including the Cox PH model and AFT models, as well as model-averaging approach, to predict how long it would take to achieve a threshold temperature of 70 °C. The findings indicate that AFT models serve as a valuable and convenient alternative to the Cox PH models, considering that they do not require the strong assumption of proportional hazards and offer a straightforward interpretation of the effects of covariates in terms of the median time, making it more intuitive and meaningful in the world of solar cookers. However, while AFT models are considered to be more advantageous over Cox PH model, they lack a straightforward way of incorporating the time-varying covariates like Cox PH models do. In addition, the model-averaging approach was utilized, indicating that it is feasible to examine the performance of the solar cooker without adhering strongly to the distribution assumptions demonstrated by the AFT models. This method is considered an innovative statistical technique for quantifying the performance of solar cooker appliances, as it has not been previously utilized within the Sc4All project context.

The findings demonstrated that the Yamo Dudo cooker has more cooking efficiency compared to Brother, Fornelia, SK14, Prototype 4 and Prototype 5, as it takes less time to attain the optimal temperature of 70 °C, according to the protocol (Funk, 2000). It is essential to reconcile the efficacy of solar cooker appliances with safety considerations. As a result, even though Yamo Dudo emerged as the better cooker, it poses substantial safety concerns. The design has a smooth parabolic surface that concentrates the entire beam at a single focal point. This intense concentration of sun rays increases the risk of serious burns or eye damage, especially among children or untrained users. The SK14 cooker, on the other hand, is parabolic but does not have a smooth surface. It has a polygonal shape with flat segments of about 10 cm, resulting in the solar beam being reflected onto a diffused plane instead of a single point. This geometric design reduces the risk of accidental exposure to highly concentrated rays, making it a safer alternative. However, although SK14 is user-friendly, it still reflects intense sunlight in flat beams, which might present a certain disadvantage in terms of health risks, posing a considerable concern for Prototype 4. These design differences underscore the necessity of heating efficiency alongside user safety, especially in real-world settings where the solar cookers may be used in unsupervised or domestic environments.

Furthermore, one of the most significant findings was the effect of using a plastic bag wrapped around the cooking pot. The findings reveal that this practice decreased the median time to attain a threshold temperature of 70 °C across all examined cooking appliances, highlighting its critical role in enhancing heat retention. This finding aligns well with prior empirical observations and theoretical expectations. The use of a plastic bag serves as an insulating layer, reducing the potential of convective and evaporative heat loss, which can significantly affect cooking efficiency. These insights enhance both the scientific comprehension and the practical advancement of solar cooking technologies.

# 5 Ethical Thinking, Societal Relevance and Stakeholder Awareness

## 5.1 Ethical Thinking

This study utilized data from experiments conducted in Belgium within the Sc4All project team at UHasselt. Data were collected for various solar cooking appliances from July 2022 to September 2024. The study emphasizes a strong commitment to transparency and scientific integrity by publicly sharing data, methodologies, and findings with the stakeholders, thereby ensuring accountability, fostering trust, and promoting collaborative efforts and awareness regarding the adoption of sustainable renewable energy sources.

## 5.2 Societal Relevance

The objective of this thesis is to facilitate the development of solar cooking devices that are affordable and work well in places where resources are limited, especially in developing countries. The study examines the integration of renewable energy into everyday cooking practices, emphasising the capacity of solar cookers to reduce reliance on conventional biomass fuels, thereby fostering environmental sustainability and improving public health. The study emphasises the need for community engagement through educational and awareness initiatives, which can help local residents in adopting solar technologies and change their energy consumption behaviours for better sustainability. This work has significant implications that align with global efforts to address climate change by reducing greenhouse effects, conserving natural resources and promoting energy independence. Therefore, the findings facilitate the pursuit of clean, reliable and affordable cooking solutions that improve people's lives and make communities that have limited resources more resilient.

### 5.3 Stakeholder Awareness

This research was the result of a collaborative endeavor involving multidisciplinary stakeholders, including UHasselt, the University of Lubumbashi in the Democratic Republic of Congo (DRC) and the Sc4All Project supported by the Flemish Government. The collaboration was instrumental in implementing an innovative approach for evaluating the performance of solar cooker appliances. The diverse expertise and insights provided by these partners enhanced the academic relevance and practical applicability of the master's thesis. Subsequently, the project team intends to travel to Tanzania to share their research findings and participate in workshops and capacity-building initiatives focused on sustainable energy technological advancements.

## 6 Recommendations for Future Experiments

To build on the findings and limitations of this thesis, as well as to deepen understanding of the solar cooker experiments within the Sc4All project, numerous recommendations can be made to guide future experimental studies. Despite the promising results obtained from the utilised survival analysis approaches, the study suffers from data management and manipulation. The experiment setup was not conducive to the use of survival methodologies, as the collected data lacked a time-to-event organization. These issues resulted in a tedious effort to develop the time-to-event dataset. Thus, in future experimental studies, it is recommended to consider the data collection approach that ensures the time-to-event structure.

The analysis used right-censored time-to-event data for convenience, but future studies should consider survival methods that properly handle the interval-censoring nature of data to improve the results. In addition, future experiments may adopt an automatic recording station developed by UH engineering students, which is currently operating in the Congo, that can record event timing on a 3-second basis, so as to improve and simplify the application of survival analysis statistical methods in this context. Moreover, to enhance the precision of the estimates, it is recommended to extend to the use of conditional survival models, such as the frailty model, that will account for the variability within different test days. This is an appropriate way since the weather might vary depending on the day the experiment is conducted, which may influence the performance of the solar cooker appliances. Furthermore, to account for the time-varying covariates in AFT models, the future studies should adopt a parsimonious approach, including manual development custom R functions, as no direct built-in function currently available for this purpose.

The study was constrained by a lack of sufficient events across several solar cooker appliances. As a result, this led to the exclusion of some cookers, such as Prototype 2 and 3, from further analysis. Thus, to address this issue, future research is recommended to undertake adequate experiments as indicated by the report on sample size determination (Mwaura, 2025). Additionally, whenever possible, future experiments should also ensure balanced data for the plastic bag usage condition to accurately detect the effect of the plastic bag and to investigate the interaction effect between a specific solar cooker and the use of a plastic bag. This will facilitate obtaining precise estimates and provide insights into the performance of a particular solar cooker while applying a plastic bag.

## References

Anderson, D. R. and Burnham, K. P. (2002). Avoiding pitfalls when using information-theoretic methods. *The Journal of wildlife management*, pages 912–918.

Bede-Ojimadu, O. and Orisakwe, O. E. (2020). Exposure to wood smoke and associated health effects in sub-saharan africa: a systematic review. *Annals of global health*, 86(1):32.

Burzykowski, T. (2024). Survival analysis: Methods for analyzing data with censored observations. In *Seminars in Orthodontics*, volume 30, pages 29–36. Elsevier.

Collett, D. (1994). Modelling survival data. In *Modelling survival data in medical research*, pages 53–106. Springer.

Cox, D. R. and Snell, E. J. (1968). A general definition of residuals. *Journal of the Royal Statistical Society: Series B (Methodological)*, 30(2):248–265.

Cuce, E. and Cuce, P. M. (2013). A comprehensive review on solar cookers. *Applied Energy*, 102:1399–1421.

DR, C. (1972). Regression models and life tables. *JR Stat Soc*, 34:248–275.

Funk, P. A. (2000). Evaluating the international standard procedure for testing solar cookers and reporting performance. *Solar Energy*, 68(1):1–7.

Gijbels, I. (2010). Censored data. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(2):178–188.

Goel, M. K., Khanna, P., and Kishore, J. (2010). Understanding survival analysis: Kaplan-meier estimate. *International journal of Ayurveda research*, 1(4):274.

Gorjian, A., Rahmati, E., Gorjian, S., Anand, A., and Jathar, L. D. (2022). A comprehensive study of research and development in concentrating solar cookers (cscs): Design considerations, recent advancements, and economics. *Solar Energy*, 245:80–107.

Grambsch, P. M. and Therneau, T. M. (1994). Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*, 81(3):515–526.

Herez, A., Ramadan, M., and Khaled, M. (2018). Review on solar cooker systems: Economic and environmental study for different lebanese scenarios. *Renewable and Sustainable Energy Reviews*, 81:421–432.

Idowu, O. S., De Azevedo, L. B., Zohoori, F. V., Kanmodi, K., and Pak, T. (2023). Health risks associated with the production and usage of charcoal: a systematic review. *BMJ open*, 13(7):e065914.

Kannan, N. and Vakeesan, D. (2016). Solar energy for future world:-a review. *Renewable and sustainable energy reviews*, 62:1092–1105.

Kauki, F. C. (2024). The use of survival analysis for the determination of the performance of solar cookers in africa.

Klein, J. P. and Moeschberger, M. L. (2003). *Censoring and Truncation*, pages 63–90. Springer New York, New York, NY.

Lagakos, S. W. (1979). General right censoring and its impact on the analysis of survival data. *Biometrics*, pages 139–156.

Liu, X. (2012). *Survival analysis: models and applications*. John Wiley & Sons.

Ma, D. (2023). *Semiparametric accelerated failure time model estimation with time-varying covariates*. PhD thesis, Macquarie University.

Mullick, S., Kandpal, T., and Saxena, A. (1987). Thermal test procedure for box-type solar cookers. *Solar energy*, 39(4):353–360.

Mwaura, P. W. (2025). Sample size determination of experiments for the determination of standardized performances of solar cooking appliances, hasselt university.

Namata, H., Aerts, M., Faes, C., and Teunis, P. (2008). Model averaging in microbial risk assessment using fractional polynomials. *Risk Analysis: An International Journal*, 28(4):891–905.

Qi, J. (2009). *Comparison of proportional hazards and accelerated failure time models*. PhD thesis, University of Saskatchewan.

R Core Team (2025). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 4.4.3 edition.

Rose, J., Bensch, G., Munyehirwe, A., and Peters, J. (2022). The forgotten coal: charcoal demand in sub-saharan africa. *World Development Perspectives*, 25:100401.

Saikia, R. and Barman, M. P. (2017). A review on accelerated failure time models. *International Journal of Statistics and Systems*, 12(2):311–322.

SAS Institute Inc. (2018). *SAS® Studio 3.8: User's Guide*. SAS Institute Inc., Cary, NC, 3.81 edition. User's Guide.

Schober, P. and Vetter, T. R. (2018). Survival analysis and interpretation of time-to-event data: the tortoise and the hare. *Anesthesia & Analgesia*, 127(3):792–798.

Şen, Z. (2004). Solar energy in progress and future research trends. *Progress in energy and combustion science*, 30(4):367–416.

Somasundaran, S. (2023). Comprehending kaplan–meier curve. *Kerala Journal of Ophthalmology*, 35(2):228–230.

Therneau, T. M., Grambsch, P. M., and Fleming, T. R. (1990). Martingale-based residuals for survival models. *Biometrika*, 77(1):147–160.

Therneau, T. M., Grambsch, P. M., Therneau, T. M., and Grambsch, P. M. (2000). *The cox model*. Springer.

Yang, S. and Prentice, R. L. (2011). Estimation of the 2-sample hazard ratio function using a semiparametric model. *Biostatistics*, 12(2):354–368.

Zhu, R., Wang, H., Zhang, X., and Liang, H. (2023). A scalable frequentist model averaging method. *Journal of Business & Economic Statistics*, 41(4):1228–1237.

Zulu, L. C. and Richardson, R. B. (2013). Charcoal, livelihoods, and poverty reduction: Evidence from sub-saharan africa. *Energy for Sustainable Development*, 17(2):127–137.

### Appendix

Table 10: *Summary results of the fitted Log-logistic AFT model with the time ratios (TR)*

| Variable | Estimate | TR | SE | p-value |
|---|---|---|---|---|
| (Intercept) | 5.1104 | 166.0179 | 0.3051 | < 0.0001 |
| Cooker2Brother | 0.4977 | 1.6450 | 0.1495 | 0.0009 |
| Cooker2Fornelia | 0.0041 | 1.0041 | 0.1834 | 0.9823 |
| Cooker2Ovenproto126 | 0.8281 | 2.2891 | 0.1786 | < 0.0001 |
| Cooker2Proto5 | 0.7172 | 2.0489 | 0.2146 | 0.0008 |
| Cooker2Sk14 | -0.3594 | 0.6982 | 0.1634 | 0.0278 |
| Cooker2YamoDudo | -0.8851 | 0.4125 | 0.1495 | < 0.0001 |
| First Temperature | -0.0238 | 0.9765 | 0.0026 | < 0.0001 |
| Average Irradiation | -0.0013 | 0.9987 | 0.0003 | < 0.0001 |
| PlasticBag | -0.1478 | 0.8625 | 0.1058 | 0.1625 |
| Cooker2Proto5:PlasticBag | -0.3758 | 0.6866 | 0.2497 | 0.1323 |

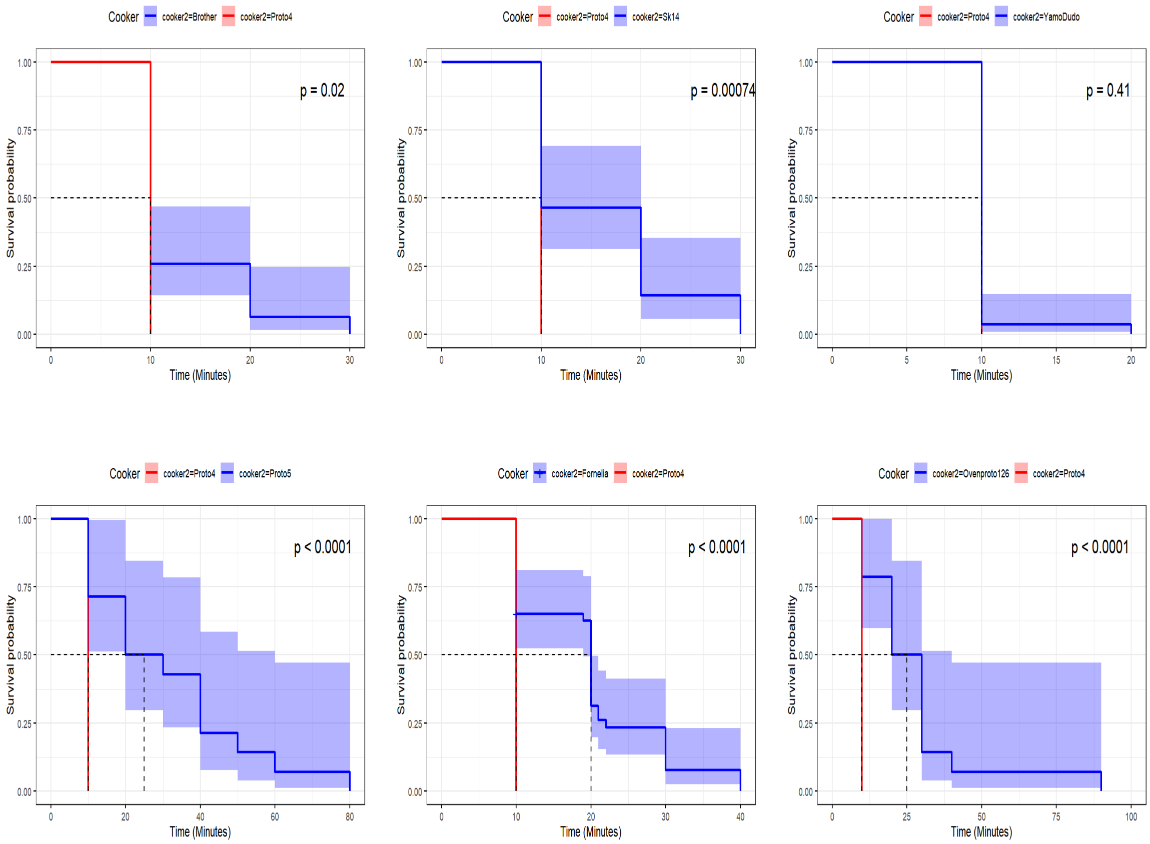


Figure 11: *Kaplan-Meier curves of the survival function for the cookers considered in the analysis. The two-sided p-values are based on a log-rank test comparing the survival functions for the cookers. The dashed lines show the medium time in minutes.*

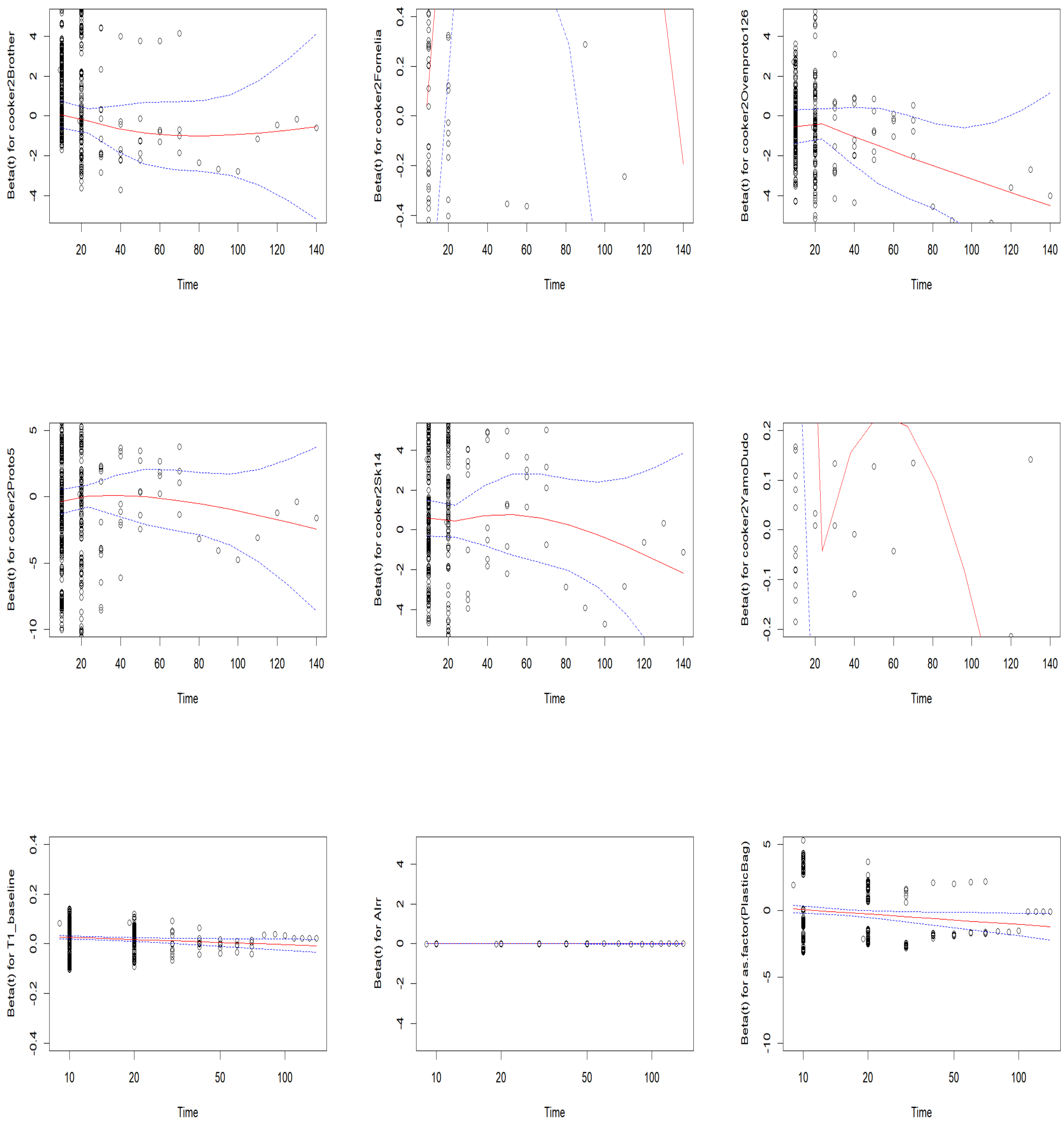


Figure 12: *Graphical diagnostics of the scaled Schoenfeld residuals for each covariate against the time. The solid red lines represent smoothing spline fits, while the dashed blue lines show standard error bands around the fit.*

### Software Code

```
# Read data
df <- read.csv("C:\\Users\\barak\\Desktop\\UHASSELT\\SECOND YEAR\\2ND SEMESTER\\MASTER
↪  THESIS\\ABOUT DATA\\DATA TO USE\\SolarCookerV20240925.csv",dec=",",sep = ";")

# A: DATA MANAGEMENT
# Transforming data to Right-censored Time to Event Data(AFT models)
# create the collapsed variable (Ovenproto126)
df$cooker2 <- dplyr::case_when(df$Cooker %in% c("OvenProto1", "OvenProto2", "Proto6") ~
↪  "Ovenproto126",TRUE ~ df$Cooker)

df$H1 <- as_hms(df$H1)
df$H2 <- as_hms(df$H2)
df$H1_temp <- factor(df$H1, "%H:%M:%S")
df$H2_temp <- factor(df$H2, "%H:%M:%S")

df$TestDate <- as.Date(df$TestDate, format = "%Y.%m.%d")

df2 <- df %>% group_by(cooker2, TestDate) %>% mutate(trial_number = 1,trial_number =
↪  ifelse((T1 == lag(T2, default = first(T2)) & H1 == lag(H2, default =first(H2))),
↪  lag(trial_number, default = 1), row_number())) %>% tidyr::fill(trial_number,
.direction = "down")

exp_window <- c(1)
for (i in 2:nrow(df2)) {
  prev_T2 <- df2$T2[i - 1]
  curr_T1 <- df2$T1[i]
if (is.na(prev_T2) || is.na(curr_T1)) {
    exp_window <- c(exp_window, exp_window[length(exp_window)] + 1)
  } else if (abs(prev_T2 - curr_T1) <= 0) {
    exp_window <- c(exp_window, exp_window[length(exp_window)])
   } else {
    exp_window <- c(exp_window, exp_window[length(exp_window)] + 1)
  }
}

df2$exp_window <- exp_window

df2_trimmed <- df2 %>% group_by(cooker2, TestDate, exp_window) %>% mutate(
first_70_idx = which(T2 >= 70)[1],row_id = row_number(), AIrr = (I1 + I2) / 2) %>%
↪  filter(is.na(first_70_idx) | row_id <= first_70_idx) %>% ungroup()

df3 <- df2_trimmed %>% group_by(cooker2, TestDate, exp_window) %>% summarise(
trial_H1 = first(H1), trial_H2 = last(H2), first_temp = first(T1), final_temp = last(T2),
↪  avg_irradiation = mean(AIrr, na.rm = TRUE),tot_time =
```

```
sum(as.numeric(difftime(last(H2), first(H1), units = "mins"))),
across(where(is.numeric) & !any_of(c("T1", "T2")), ~mean(.x, na.rm = TRUE)),
↪  across(where(is.character), first),
.groups = "drop")

df4 <- df3 %>% mutate(Event = ifelse(final_temp >= 70, 1, 0))
df5$cooker2 <- trimws(as.character(df4$cooker2))
selected_cookers <- c("Proto4","Proto5","YamoDudo", "Brother","Fornelia",
"Sk14","Ovenproto126")
df_subset <- df5[df5$cooker2 %in% selected_cookers, ]
df_subset$cooker2 <- factor(df_subset$cooker2)

# B: DATA EXPLORATORY
# Table of summary
summary_table <- df_subset %>%group_by(cooker2, Event) %>% summarise(count = n(), .groups
↪  = "drop") %>%tidyr::pivot_wider(names_from = Event, values_from = count, values_fill
↪  = list(count = 0)) %>% rename(Censor = `0`, Event = `1`) %>% mutate(Total = Censor +
↪  Event)
print(summary_table)

summary_table2 <- df_subset %>%group_by(Cooker, PlasticBag) %>% summarise(count = n(),
↪  .groups = "drop") %>%tidyr::pivot_wider(names_from = PlasticBag, values_from = count,
↪  values_fill = list(count = 0)) %>%rename(No = `0`, Yes = `1`) %>% mutate(Total = No +
↪  Yes)
print(summary_table2)

# Kaplan-Meier curves
fit <- survfit(Surv(tot_time, Event) ~ Cooker2, data = df_subset)
ggsurv3 <- ggsurvplot(fit,legend.title = "Cooker",xlab = "Time (Minutes)",
risk.table = FALSE,conf.int = TRUE,conf.type = "log-log",tables.height = 0.3,pval =
↪  TRUE,pval.coord = c(25, 0.90), surv.median.line = "hv", ggtheme = theme_bw(),
palette = c("red", "blue") )
print(ggsurv3)

# C: ANALYSIS
# Fit AFT models
(Generalized_Gamma <- flexsurvreg(Surv(tot_time, Event) ~  cooker2Brother
↪  +cooker2Fornelia + cooker2Ovenproto126 + cooker2Proto5 + cooker2Sk14 +
↪  cooker2YamoDudo + first_temp + avg_irradiation + PlasticBag,data = df_subset,
↪  dist="gengamma"))

(Generalized_F <- flexsurvreg(Surv(tot_time, Event) ~  cooker2Brother +cooker2Fornelia +
↪  cooker2Ovenproto126 + cooker2Proto5 + cooker2Sk14 + cooker2YamoDudo + first_temp +
↪  avg_irradiation + PlasticBag,data = df_subset, dist="genf"))
```

```
(lognormal <- survreg(Surv(tot_time, Event) ~ cooker2Brother +cooker2Fornelia +
↪  cooker2Ovenproto126 + cooker2Proto5 + cooker2Sk14 + cooker2YamoDudo + first_temp +
↪  avg_irradiation + PlasticBag,data = df_subset, dist="lognormal"))

(loglogistic <-  survreg(Surv(tot_time, Event) ~ cooker2Brother +cooker2Fornelia +
↪  cooker2Ovenproto126 + cooker2Proto5 + cooker2Sk14 + cooker2YamoDudo + first_temp +
↪  avg_irradiation + PlasticBag,data = df_subset, dist="loglogistic"))

(weibull <-  survreg(Surv(tot_time, Event) ~ cooker2Brother +cooker2Fornelia +
↪  cooker2Ovenproto126 + cooker2Proto5 + cooker2Sk14 + cooker2YamoDudo + first_temp +
↪  avg_irradiation + PlasticBag,data = df_subset, dist="weibull"))

# AIC and BIC values for each model
extractAIC(model)[2]
BIC(model)

# Pairwise comparisons (contrasts)
emm_pb1 <- emmeans(loglogistic, ~ cooker2)
pairwise_contrasts <- contrast(emm_pb1, method = "pairwise")
(emm<- summary(pairwise_contrasts, adjust = "holm"))

# Assessing goodness of fit
LL <- psm(Surv(tot_time, Event) ~ cooker2Brother +cooker2Fornelia + cooker2Ovenproto126 +
↪  cooker2Proto5 + cooker2Sk14 + cooker2YamoDudo + first_temp + avg_irradiation +
↪  PlasticBag,data = df_subset, dist="loglogistic", y=TRUE)
res.LL <-resid(LL,type="cens")
survplot(npsurv(res.LL ~1),conf ="none",ylab="Loglogistic", xlab="Residual")
lines(res.LL, lwd = 0.5, col = "red")

### cox-snell residual plot (SAS CODE)
/* Step 2: Fit the model */
proc lifereg data=WORK.IMPORT_CLEAN1;
    class cooker2Brother cooker2Fornelia cooker2Ovenproto126 cooker2Proto5 cooker2Sk14
    ↪  cooker2YamoDudo PlasticBag;
model tot_time*Event(0) = cooker2Brother cooker2Fornelia cooker2Ovenproto126
↪  cooker2Proto5 cooker2Sk14 cooker2YamoDudo PlasticBag first_temp avg_irradiation/
↪  dist=llogistic;
output out=resids sres=stdres cres=coxsnell;run;

proc lifetest data=resids outsurv=surv_cs;
time coxsnell*Event(0); run;

data surv_cs;
```

```
set surv_cs;
lsurv=-log(survival);
expct=coxsnell; run;

proc gplot data=surv_cs;
axis1 label=(f='Arial' h=1.8 'Cox-Snell Residual') value=(f='Arial' h=2.8) order=0 to 6
↪  by 1 style=1 width=1 color=black;
axis2 label=(angle=90 f='Arial' h=1.8 '-Ln(KM survival)') value=(f='Arial' h=2.8)
order=0 to 6 by 1 style=1 width=1 color=black;
symbol1 interpol=none value=dot h=1 c=black;
symbol2 interpol=join value=none l=1 c=red w=1;
plot lsurv*coxsnell expct*coxsnell / overlay vaxis=axis2 haxis=axis1;
run; quit;

# Model-averaging approach
model.set <- list(m1, m2, m3)
names(model.set) <- c("lognormal", "weibull", "loglogistic")
model.sel <- model.sel(model.set)
print(model.sel)
avg.model <- model.avg(model.set)
summary(avg.model)

# Prediction of the Median Survival Time
# Define coefficients
coefs <- list(intercept = 5.1855,CookerBrother = 0.4951,CookerFornelia =
↪  -0.0571,CookerOvenproto126 = 0.7952,CookerProto5 = 0.5346,CookerSk14 = -0.4002,
CookerYamoDudo = -0.9392,temp = -0.0239,irradiance = -0.0013,PlasticBag =
↪  -0.2109,CookerProto4 = 0 )

# Define standard errors for asymptotic prediction intervals
coefs_se <- list(intercept = 0.3027,CookerBrother = 0.1499,CookerFornelia = 0.1794,
↪  CookerOvenproto126 = 0.1785, CookerProto5 = 0.1772, CookerSk14 =
↪  0.1614,CookerYamoDudo = 0.1454,temp = 0.0026,irradiance = 0.0003,PlasticBag =
↪  0.0985,CookerProto4 = 0 )

# conditions
baseline_temp <- 20
avg_irradiance <- 700

# Helper function for median prediction with PI
predict_median_with_PI <- function(cooker, plastic_bag) {cooker_coef <- ifelse(cooker
↪  %in% names(coefs), coefs[[cooker]], 0)cooker_se <- ifelse(cooker %in%
↪  names(coefs_se), coefs_se[[cooker]], 0)pb_coef <- ifelse(plastic_bag,
↪  coefs$PlasticBag, 0)pb_se <- ifelse(plastic_bag, coefs_se$PlasticBag, 0)
```

```
log_time <- coefs$intercept + cooker_coef +coefs$temp * baseline_temp + coefs$irradiance
↪  * avg_irradiance +pb_coef

log_time_se <- sqrt(coefs_se$intercept^2 + cooker_se^2 + pb_se^2)

median_time <- exp(log_time)
lower_PI <- exp(log_time - 1.96 * log_time_se)
upper_PI <- exp(log_time + 1.96 * log_time_se)

return(c(median_time, lower_PI, upper_PI))}

# List of cookers
cookers <- c("CookerProto4", "CookerBrother", "CookerFornelia",
↪  "CookerOvenproto126","CookerProto5", "CookerSk14", "CookerYamoDudo")

# Compute predictions with PI
predictions <- data.frame(Cooker = cookers,No_Bag = t(sapply(cookers, function(x)
↪  predict_median_with_PI(x, FALSE))),With_Bag = t(sapply(cookers, function(x)
↪  predict_median_with_PI(x, TRUE))))

# Rename columns for clarity
colnames(predictions) <- c("Cooker", "No_Bag_Median", "No_Bag_Lower_PI",
↪  "No_Bag_Upper_PI","With_Bag_Median", "With_Bag_Lower_PI", "With_Bag_Upper_PI")

# Print final dataframe
print(predictions)
```