2016

The Reflected-Shifted-Truncated-Gamma Distribution for Negatively Skewed Survival Data with Application to Pediatric Nephrotic Syndrome

Sophia D. Waymyers
University of South Carolina

The Reflected-Shifted-Truncated-Gamma Distribution for Negatively Skewed Survival Data with Application to Pediatric Nephrotic Syndrome

by

Sophia D. Waymyers

Bachelor of Science
Winthrop University 1990

Master of Science
University of South Carolina 1998

Master of Science
University of South Carolina 2011

Submitted in Partial Fulfillment of the Requirements

for the Degree of Doctor of Philosophy in

Biostatistics

The Norman J. Arnold School of Public Health

University of South Carolina

2016

Accepted by:

Hrishikesh Chakraborty, Major Professor

Alexander McLain, Committee Member

Christine Turley, Committee Member

Sanku Dey, Committee Member

Cheryl L. Addy, Vice Provost and Dean of the Graduate School

Dedication

I dedicate this dissertation to my husband Anthony for always believing in me, and to the four greatest accomplishments of my life : my beautiful daughters, Kayla, Ashley, and Raven, and my son, Anthony II, who continues to rise above the challenges presented by the idiopathic pediatric nephrotic syndrome.

Acknowledgments

The completion of this dissertation has only been possible through the grace of God, the guidance and assistance of committee members, and the support of family and friends.

I would like to express my most sincere gratitude to Dr. Hrishikesh Chakraborty (Dr. Rishi), my dissertation advisor, for guiding me from a basic idea to this completed product. His insight, suggestions, and new ideas have been invaluable, as well as his availability and tireless efforts to encourage me to improve. I thank my committee members for their time and patience in this process: Dr. Alexander McLain for his timely remarks and guidance, especially during the final stages of my graduate studies, Dr. Christine Turley for her support and medical expertise on the nature of atopy and nephrotic syndrome, and Dr. Sanku Dey for inspiring me with timely and careful responses to my work despite the eight and one-half hour difference in time zone. To Dr. Don Edwards, Dr. Edsel Pena, Dr. James Hussey and Dr. Robert Moran, I say thank you for the role you played in helping a dream come true. I am grateful to the many other faculty, staff and graduate students at USC that I have encountered during my tenure as a graduate student. I have indeed been humbled by the entire experience, and I have enjoyed the journey.

To my daughters and my friends who offered encouragement on my journey when I wanted to give up, I again say thank you. Thank you for laughing with me, crying with me, and celebrating with me. I am especially grateful for the two friends that read rough drafts of my work and used their superb editorial skills to dramatically improve my writing. I thank my colleagues in the Department of Mathematics and

the administration at Francis Marion University for being supportive of my goals. Finally, I thank my parents, Benjamin and the late Claretta Funchess, for allowing me to dream.

.

Abstract

Negatively skewed survival data arise occasionally in public health fields and in statistical research. Standard distributions such as the exponential, generalized F, generalized gamma, Gompertz, log-logistic, lognormal, Rayleigh, and Weibull distributions are not always well suited to this data. The primary goal of this dissertation is to find a viable alternative for modeling negatively skewed survival data such as the time to first remission for pediatric patients with frequently relapsing or steroid dependent nephrotic syndrome.

We begin with a brief introduction of survival analysis and the nature of pediatric nephrotic syndrome. A meta-analysis on atopy and pediatric nephrotic syndrome using worldwide studies is performed. We introduce the reflected-shifted-truncated-gamma (RSTG) distribution as an alternative model for survival data whose event times arise from a negatively skewed distribution. Explicit expressions are provided for the mean, variance, hazard function, survival function and quantile function of the RSTG distribution. A simulation study verifies the consistency of maximum likelihood estimates of model parameters. Using maximum likelihood methods, we compare the RSTG distribution to the exponential, generalized F, generalized gamma, Gompertz, log-logistic, lognormal, Rayleigh, and Weibull distributions for modeling negatively skewed complete (uncensored) data, right-censored data and interval-censored data using well-known data sets. We then apply the RSTG distribution to pediatric nephrotic syndrome data from the Clinical Data Warehouse from Health Sciences of South Carolina and from the Robert Wood Johnson Medical School in New Jersey using covariate adjusted accelerated failure time (AFT) models with and without

frailty. We include a brief example of the RSTG distribution applied to a 1972 study on diabetic retinopathy.

Our research shows that the RSTG distribution is superior to the eight aforementioned distributions for modeling negatively skewed survival data. The results from applications of this distribution and future goals are discussed.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AFT ..... Accelerated Failure Time

AIC ..... Akaike Information Criterion

AICC ..... corrected Akaike Information Criterion

CDW ..... Clinical Data Warehouse

DRS ..... Diabetic Retinopathy Study

EM ..... Expectation-Maximization

FRNS ..... Frequently Relapsing Nephrotic Syndrome

HLA ..... Human Leukocyte Antigen

HQIC ..... Hannan-Quinn Information Criterion

HSSC ..... Health Sciences of South Carolina

IFRNS ..... Infrequently Relapsing Nephrotic Syndrome

ISKDC ..... International Study of Kidney Disease in Children

KDIGO ..... Kidney Disease-Improving Global Outcomes

MCD ..... Minimal Change Disease

NS ..... Nephrotic Syndrome

PH ..... Proportional Hazards

RSTG ..... Reflected-Shifted-Truncated-Gamma

SDNS ..... Steroid Dependent Nephrotic Syndrome

SRNS ..... Steroid Resistant Nephrotic Syndrome

S.R.N.S. ..... Steroid Responsive Nephrotic Syndrome

Chapter 1

Background and Significance

1.1 Survival Analysis

Survival analysis is a statistical method for data analysis in which the outcome variable is the time to the occurrence of an event (John P. Klein and Moeschberger, 2003). Time-to-event data, or survival time data, is common in medical research. Examples of events of interest include relapse of cancer, remission of nephrotic syndrome, recurrence of a tumor, development time from HIV infection to an AIDS diagnosis for HIV patients, or death. The definition of event time should be made clear at the start of the study. In studying the nature of the disease, for example, we must specify whether the time of origin is when the symptoms start, when the biological identification of the disease happens, or when the diagnosis is made. The time scale and the origin of the event must also be identified.

Survival times can either be censored or uncensored. Uncensored observations are commonly referred to as complete data and are observed exactly. Censored observations are not observed exactly and can be left-censored, right-censored, or interval-censored. If the event of interest occurs prior to the start of a study, the observation time is left-censored. An observation is right-censored if a subject withdraws from a study or if the study ends before the event of interest has occurred. An observation is interval-censored if the event of interest is only known to have occurred within a given interval of time. In this dissertation, we will assume that all censoring is non-informative, or unrelated to the study.

Survival analysis techniques make use of the incomplete data collected from censored event times. The event times are subject to random variation, and, like any random variables, form a distribution (Lee and Wang, 2003). The distribution of survival times is usually characterized by three functions: (1) the survival function, (2) the probability density function, and (3) the hazard function. These three functions are mathematically equivalent. If one of them is known, the other two can be derived.

1.1.1 Basics of Survival Analysis

A function $f(u)$ is a **probability density function** (pdf) of a random variable U if and only if

1. $f(u) \geq 0$ for all u
2. $\int_{-\infty}^{\infty} f(u)\mathrm{d}u = 1.$

The **cumulative distribution function**, or cdf, of a random variable U, denoted by $F(u)$, is defined by

$$\begin{aligned} F(u) &= P(U \leq u), \\ &= \int_{-\infty}^{u} f(x)\mathrm{d}x. \end{aligned}$$

The area under the curve of $f(u)$ can give interval probabilities. For example, $P(a < U < b) = \int_{a}^{b} f(u)\mathrm{d}u = F(b) - F(a)$. If $f(u)$ is continuous, the Fundamental Theorem of Calculus gives the additional relationship

$$\frac{d}{du}F(u) = f(u).$$

The **survival function** gives the probability that a person survives longer than some specified time t. Let T denote a nonnegative random variable whose individual values t represent the time to an event of interest. The survival function of T, denoted by $S(t)$, is given by $S(t) = P(T > t), 0 < t < \infty$. Theoretically, the survival

function is a non-increasing continuous function such that $S(0) = 1$ and $S(\infty) = 0$. If the random variable T has some underlying probability density function $f(u)$, then $S(t) = \int_t^\infty f(u)\mathrm{d}u$, $0 < t < \infty$. Note that $S(t) = 1 - F(t)$.

The **hazard function** gives the instantaneous potential per unit time for the event to occur, given that the event has not occurred up to time t. It is given by

$$\begin{aligned}
h(t) &= \lim_{\Delta t \to 0+} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} \\
&= \lim_{\Delta t \to 0+} \frac{\dfrac{P(t \leq T < t + \Delta t) \cap P(T \geq t)}{P(T \geq t)}}{\Delta t} \\
&= \lim_{\Delta t \to 0+} \frac{P(t \leq T < t + \Delta t)}{\Delta t} \cdot \frac{1}{P(T \geq t)} \\
&= \lim_{\Delta t \to 0+} \frac{F(t + \Delta t) - F(t)}{\Delta t} \cdot \frac{1}{P(T \geq t)} \\
&= \frac{d}{dt} F(t) \cdot \frac{1}{P(T \geq t)} \\
&= \frac{f(t)}{S(t)},
\end{aligned}$$

where $0 < t < \infty$. The hazard function is sometimes called a conditional failure rate (Kleinbaum and Klein, 2006).

The cumulative hazard function is defined as $H(t) = P(T \leq t) = \int_0^t h(u)\mathrm{d}u$, $0 < t < \infty$,

and can be derived from the density and survival functions as follows:

$$\begin{aligned}
h(t) &= \frac{f(t)}{S(t)} \\
&= \frac{\dfrac{d}{dt} F(t)}{S(t)} \\
&= \frac{\dfrac{d}{dt}(1 - S(t))}{S(t)} \\
&= \frac{-\dfrac{d}{dt}(S(t))}{S(t)}.
\end{aligned}$$

Therefore, we can express $H(t) = \int_0^t h(u)\mathrm{d}u = -\log S(t)$.

1.1.2 Goals of Survival Analysis

The goals of survival analysis include estimation of the survival and/or hazard functions, comparison of the survival and/or hazard functions, and assessment of the relationship between explanatory variables (covariates) and survival time. Survival data consists of a time of event, generally denoted by t_i, the values of any covariates that are considered valuable to the model, generally denoted by x_i, and a censoring indicator, δ_i, in which

$$\delta_i = \begin{cases} 1 & \text{if the event of interest is observed} \\ 0 & \text{if the event of interest is not observed} \end{cases}.$$

One of the first steps in analyzing survival data is presenting a numerical or graphical summary of the data. This summary may be used to suggest a survival model for the data. Non-parametric models are widely used models in the literature. The most common of these, the **Kaplan-Meier estimator** of the survival function, is actually a step function in which the estimated survival probabilities are constant between adjacent event times and decrease with each event time. If there is no censoring, the function is simply a step function equal to the proportion surviving an instant after time t. One common use of the Kaplan-Meier estimated survival function is to compare the survival probabilities of two groups. A major disadvantage of this type of model is its inability to estimate survival probabilities at all time points.

The **Cox proportional hazards (PH) model** is most common for modeling survival data to assess the effect of covariates on survival probabilities. The Cox model is often deemed as semi-parametric because no assumptions are made concerning a baseline hazard function; consequently, no assumptions are made on the distribution of the survival times. It is assumed that all subjects in the study have a common baseline hazard; thus, the ratio of the hazards of two subjects is some constant that is

independent of time. The model is specified as $h(t) = h_0(t)e^{\beta x}$, where $0 < t < \infty$, $x = (x_1, x_2, \ldots, x_n)$ represents the vector of covariates of interest, e^{β} represents the hazard ratio between groups and $h_0(t)$ represents the baseline hazard function. An advantage of this model is its considerable flexibility since full specification of the baseline hazard function is not required. Fully parametric proportional hazards models, which assume that the baseline hazard function can be fully parametrized, can also be used. A disadvantage of both the semi-parametric and the fully parametric proportional hazards models is the necessity of the assumption of proportional hazards, which may not be valid for some time-to-event data. Alternative versions of the standard Cox proportional hazards model can be used when the proportional hazards assumption is not satisfied. These include the **time dependent Cox model** and the **stratified Cox model**. The time dependent Cox model introduces a time dependent variable to accommodate the variable that does not satisfy the proportional hazards assumption. The stratified Cox model allows the model to be stratified on the variable that does not satisfy the proportional hazards assumption. Either model can limit the overall interpretability of the parameters as the variable that has been modified to accommodate the proportional hazards assumption is no longer interpretable.

The **accelerated failure time (AFT) model** is a general parametric model for survival data that has been used more frequently in recent years. In the accelerated failure time model, covariates have a direct effect on the survival time while in the proportional hazards model, the covariates have a multiplicative effect on the hazard function. The covariate effects change the timescale in the AFT model and, therefore, accelerate or decelerate the time to the event of interest. The general model for accelerated failure time is $S(t) = S_0(\gamma t)$ where $S_0(t)$ represents a baseline survival function and γ represents an acceleration factor. Parametric AFT models provide a useful alternative to the PH model when modeling survival data (Wei, 1992). The AFT approach models the survival times directly and gives summary measures that

are interpreted in terms of the survival curve (Hutton and Monaghan, 2002; Orbe, Ferreira, and Núñez-Antón, 2002; Patel, Kay, and Rowell, 2006; Pourhoseingholi et al., 2007). Common applications of the parametric AFT model in the literature include aging research (Swindell, 2009), kidney transplant survival (Lambert, Collett, Kimber, and Johnson, 2004), and coronary heart disease (Chen, Zhang, and Zhang, 2013). The parametric AFT model can incorporate a wide range of survival distributions.

One of the main advantages of using parametric survival models is the complete specification of the survival, hazard, and density functions. Also, the parametric model is smooth and continuous. The parametric approach can estimate "between point" probabilities whereas the non-parametric approach can only give stepwise-estimates using the time points actually reported in the study. Nevertheless, parametric survival models are historically not as popular as non-parametric or semi-parametric models. Derivations of the survival and hazard functions of parametric models may be computationally intensive, and the true nature of a distribution may be hard to verify in practice. However, it is generally agreed that if a parametric form can capture the true nature of data, the results and implications will be far more precise (Collett, 2015).

1.2 Pediatric Nephrotic Syndrome

"Idiopathic" nephrotic syndrome, or nephrotic syndrome that arises spontaneously, is a rare disease syndrome that commonly has a relapsing course. Pediatric idiopathic nephrotic syndrome, a condition listed in the Rare Diseases Clinical Research Network, a division of the National Institute of Health, is a condition that about 2-7 of every 100,000 children are living with today (Kerlin, Haworth, and Smoyer, 2014). According to the National Institute of Diabetes and Digestive and Kidney Diseases, pediatric nephrotic syndrome is a set of signs or symptoms that may indicate kidney dysfunction. The nephrotic syndrome describes a triad of hypoalbuminemia (low levels

of protein in the blood), edema (swelling resulting from buildup of salt and water), and hyperlipidemia (high levels of protein in the urine)(Saleem, 2013). Other signs include less frequent urination and weight gain from excess fluid. Pediatric nephrotic syndrome can occur at any age but most commonly occurs between the ages of $1\frac{1}{2}$ and 5 years of age and affects boys more than girls (Childhood Nephrotic Syndrome, 2016).

Prior to the initiation of steroid treatment in the 1960s, the risk of morbidity and mortality from pediatric nephrotic syndrome was extremely high (Soyka, 1967). It is now the widely accepted standard that the best first line treatment for the initial diagnosis of idiopathic nephrotic syndrome in children is a high dosage corticosteroid treatment (Fomina, Pavlenko, Englund, and Bagdasarova, 2011; Noer, 2005; Richardson, 2012; Pasini et al., 2015). However, prolonged and repeated corticosteroid treatment may induce serious steroid toxicity such as growth retardation and cataracts (A. Takeda, Matsutani, Niimura, and Ohgushi, 1996). Other significant side effects of these treatments include high blood pressure, increased appetite and significant weight gain, restlessness, behavioral changes, reduction in the body's ability to fight infection, cosmetic side effects such as increased hair growth on the face or body, swollen or painful gums, and, less commonly, painful urination (CVS Pharmacy, 2016).

In general, if a relapse occurs several times within a given time frame, the diagnosis of nephrotic syndrome is further classified as either steroid dependent (sometimes referred to in earlier literature as steroid responsive) or frequently relapsing. In this case, a second line of medication is introduced to reduce the risk of steroid toxicity and to achieve a lasting remission. If a first remission is still not achieved within a given time frame, the diagnosis becomes steroid resistant, and alternative treatment methods are applied (Lombel, Gipson, and Hodson, 2013). Definitions of the common classifications of pediatric nephrotic syndrome are given in Appendix A. Previous findings suggest that the majority of pediatric patients will relapse after the initial

remission. Thus, other medicines are studied to determine their effectiveness in achieving and maintaining a remission for as long as possible (Fomina et al., 2011; Mishra, Abhinay, Mishra, Prasad, and Pohl, 2013; A. Takeda et al., 1996). Current treatment strategies in addition to the high dose of prolonged steroid treatment include cyclophosphamide, chlorambucil, cyclosporine A, mycophenolate mofetil and other immunosuppressive agents (Fomina et al., 2011). More recently, rituximab, an intravenous drug used to treat rheumatoid arthritis and B-cell non-Hodgkin's lymphoma, has been used and is under study for the difficult-to-treat nephrotic syndrome (Sinha et al., 2015).

Studies suggest that almost all proposed and currently used treatments carry significant side effects (Latta, von Schnakenburg, and Ehrich, 2001; Iijima et al., 2002; Tullus and Marks, 2013). Researchers urgently need to better understand the nature of the disease and to identify specific risk factors that would foster better treatment decisions. The pediatric patient diagnosed with frequently relapsing nephrotic syndrome or steroid dependent nephrotic syndrome provides a special challenge to the parent and health care provider since the risk of adverse events from prolonged or repeated medication is much higher.

1.2.1 Previous Findings and Methods

Much of the literature involving pediatric idiopathic nephrotic syndrome originates in areas other than the U.S. In a recently documented multicenter retrospective study, six pediatric nephrology units in Italy collected data and studied the regimens for management of the disease (Pasini et al., 2015). This study highlights the vast differences in treatment strategies and efforts to prevent acute complications from the disease, while shedding light on many of the epidemiological, clinical, and laboratory parameters of pediatric patients diagnosed with idiopathic nephrotic syndrome. Studies have been conducted in Bangladesh (Sarker et al., 2012) with more concentration on

age, socioeconomic status and rate of infection. A study of immunosuppressive agents in pediatric nephrotic syndrome was also conducted in Australia (Durkan, Hodson, Willis, and Craig, 2001).

Sixteen institutions in North America conducted a large cross sectional study on patient reported outcomes with change in nephrotic syndrome relapse or remission status (D. S. Gipson et al., 2013). According to a division of the U.S. National Institute of Health, there are also several active or recruiting clinical trials for research on pediatric nephrotic syndrome (Pediatric Nephrotic Syndrome, 2016). One of those is a large scale observational cohort study known as INSIGHT (Insight into Nephrotic Syndrome: Investigating Genes, Health and Therapeutics). It is currently studying and recruiting patients in Canada (Hussain et al., 2013). Only five of the studies listed by the U.S. National Institute of Health involve U.S data, and only one has been completed with published results.

Literature suggests that relapses within the first year of diagnosis are highly predictive of the subsequent course of the disease. This finding was confirmed in India (Mishra et al., 2013), Japan (A. Takeda et al., 1996), China (Wang, Liu, Dai, Yang, and Tang, 2005), Indonesia (Noer, 2005) and the Ukraine (Fomina et al., 2011). Other factors, including gender, age at onset, and the tapering regimen for steroid therapy, were found to be insignificant in predicting subsequent relapse. Numerous reports have suggested an association between atopy and nephrotic syndrome (Thomson, Stokes, Barratt, Turner, and Soothill, 1976; Meadow and Sarsfield, 1981; Rebien, Müller-Wiefel, Wahn, and Schärer, 1981; Yap et al., 1983; Hilmanto, 2007; Abdel-Hafez, Shimada, Lee, Johnson, and Garin, 2009). Past analysis efforts for pediatric nephrotic syndrome have included basic univariate analyses, logistic regression, use of the Kaplan-Meier survival model, and use of the Cox proportional hazards model. These measures were taken to describe the overall characteristics of the patient, to assess the relative contribution of factors affecting the relapse status of the patient,

and to analyze the efficacy of various treatment strategies (ISKDC; Tarshish, Tobin, Bernstein, and Edelmann, 1997; Constantinescu, Shah, Foote, and Weiss, 2000; Wang et al., 2005; Debbie S. Gipson et al., 2009; Fomina et al., 2011; Ishikura et al., 2012). Studies also assess predictors for and frequency of relapse (A. Takeda et al., 1996; Fomina et al., 2011; Mishra et al., 2013; Sureshkumar, Hodson, Willis, Barzi, and Craig, 2014). Gadegbeku et al. (2013) state that the ability to effectively treat nephrotic syndrome is hindered by a lack of understanding of disease mechanisms and lack of predictors to identify clinical course and therapeutic responsiveness.

For the pediatric patient whose diagnosis is frequently relapsing or steroid dependent nephrotic syndrome, studies suggest that the time to initial remission is significantly longer than those diagnosed with other, more manageable forms of the disease, such as infrequently relapsing or non-relapsing nephrotic syndrome (Vivarelli, Moscaritolo, Tsalkidis, Massella, and Emma, 2010; Yap, Han, Heng, and Gong, 2001; Letavernier et al., 2008; Fujinaga, Hirano, and Nishizaki, 2011; Nakanishi et al., 2013; Constantinescu et al., 2000; Harambat et al., 2013; Sureshkumar et al., 2014). No prior studies address the possible effect of a covariate to accelerate or decelerate the time to first remission. None of the literature to date has analyzed predictors for remission using the accelerated failure time model. Furthermore, since there are distinct geographical, economic, technological, and cultural differences between the U.S. and other regions, researchers need to perform more studies on the U.S. population.

1.3 Pediatric Nephrotic Syndrome Data Sources

1.3.1 South Carolina

Health Sciences of South Carolina (HSSC), the first statewide biomedical research collaborative in the United States, has established a database that includes data on pediatric nephrotic syndrome (Research, 2016). This statewide Clinical Data

Warehouse (CDW) system is a part of its mission to improve the health of all South Carolinians. The creation of the CDW and the data management platform support the goal of significant growth in clinical trials and medical research by facilitating collaboration across HSSC member organizations (Clinical Data Warehouse, 2016). The data include demographics, visits/encounters, diagnoses, procedures, labs, and medications. The database will contain 3.2 million patients of all ages with various ailments and diseases across South Carolina. It includes longitudinal data files with real time updates. Current data is reflective of 2004-2015. We obtained data on pediatric nephrotic syndrome patients from this database for use in our analysis. All permissions were obtained for data access and use.

1.3.2 New Jersey

The Robert Wood Johnson Medical School in New Brunswick, N.J. is one of the nation's leading comprehensive medical schools. Previously an academic unit of the University of Medicine and Dentistry of New Jersey, Robert Wood Johnson Medical School transferred to Rutgers University as part of the New Jersey Medical and Health Sciences Education Restructuring Act, on July 1, 2013 (About RWJMS, 2016). Pediatric nephrologists from Robert Wood Johnson Medical School in New Jersey performed a retrospective chart review of all pediatric patients with nephrotic syndrome that were followed up for at least one year (Constantinescu et al., 2000). The data collected at the initial diagnosis of NS included gender, race, age, hematuria status, days to remission, and pattern of relapses in the first year after diagnosis. Data necessary for the analysis were obtained from the study authors.

1.4 Dissertation Goals

In this dissertation, we concentrate on survival methods for modeling negatively skewed data. Pediatric nephrotic syndrome is used as a motivating example for our

research. We begin in Chapter 2 by assessing the relationship between atopy and nephrotic syndrome using a meta-analysis of worldwide studies. We provide a brief descriptive analysis of both the South Carolina pediatric nephrotic syndrome data and the New Jersey data in Chapter 3. In Chapter 4, we develop the reflected-shifted-truncated-gamma (RSTG) distribution for use in modeling negatively skewed data. Also in Chapter 4, we provide explicit expressions for the mean, variance, hazard function, survival function and quantile function of the RSTG distribution. We estimate the model parameters by maximum likelihood methods based on complete, right-censored and interval-censored survival data. We assess the performance and verify the consistency of the maximum likelihood estimators of the RSTG distribution by conducting a simulation study with various sample sizes, and we compare the RSTG distribution to the exponential, generalized F, generalized gamma, Gompertz, log-logistic, lognormal, Rayleigh, and Weibull distributions when modeling negatively skewed data in three real data sets.

In Chapter 5, we use the RSTG distribution in an accelerated failure time model, apply it to the pediatric nephrotic syndrome data and draw conclusions. We use the RSTG distribution in an accelerated failure time model with frailty in Chapter 6. Finally, we summarize our research and discuss future work in Chapter 7.

Chapter 2

Atopy and Pediatric Nephrotic Syndrome: A Meta-Analysis

2.1 Introduction

Pediatric nephrotic syndrome has been sporadically studied in many different countries for over 30 years. An association between nephrotic syndrome and atopic activity has been noted in multiple studies (studies 4-23 in Appendix B), but identification of allergies as a specific risk factor was not always a primary goal of the study. Furthermore, the definition of allergy/atopy was not standardized and the wide heterogeneity in terminology used to define allergy could make results unclear.

Allergies are a common public-health concern. A proclivity to allergies may cause an individual's immune system to operate in a more heightened state than normal. Natural mechanisms of the body that fight foreign antigens produce antibodies that may react in other places in the human body. These reactions could cause adverse effects. For example, the antibodies could bind to membranes in the kidneys, causing damage and leakage that potentially leads to kidney disorders such as nephrotic syndrome (National Institute of Health, 2014). Moreover, medications commonly used to treat nephrotic syndrome work to suppress the immune system and might inadvertently suppress the ability of the body to police the role of the antibodies (National Institute of Health, 2014). These medications, such as prednisone, chlorambucil, and cyclosporine, may be nephrotoxic themselves.

The connection between the pediatric nephrotic syndrome and atopy, or more generally the immune system, requires further study. This study quantifies the association between atopy and nephrotic syndrome by analyzing previous studies on atopic activity and the presence or absence of nephrotic syndrome in pediatric patients.

2.2 Methods

2.2.1 Literature Search

Published reports involving pediatric nephrotic syndrome and atopy were acquired from searches conducted from February 2014 to June 2014. The searches were conducted using NIH (National Institutes of Health) registry of studies, the Cochrane Collaboration, PubMed and PubMed Central, Embase, Google Scholar, Medline, CDSR (Cochran Database of Systematic Reviews), NICE (National Institute for Health and Care Excellence), Medscape and ProQuest. The following medical subject headings and terms were used: nephrotic syndrome, pediatric nephrotic syndrome, allergy, and atopy. Other sources were found in the references section of the retrieved articles and from two pediatric nephrologists known to be actively involved with pediatric nephrotic syndrome. 173 publications were obtained. No location, language or time restrictions were applied.

2.2.2 Study Selection

Any study article that referenced a relationship between nephrotic syndrome in the pediatric population and atopy was included in the first phase of study selection. The pediatric population was limited to individuals between 0 and 18 years of age. From the 20 studies selected in this phase, we excluded studies without adequate information to calculate an odds ratio and corresponding 95% confidence interval, studies that selected controls with regard to exposure status, and studies that included

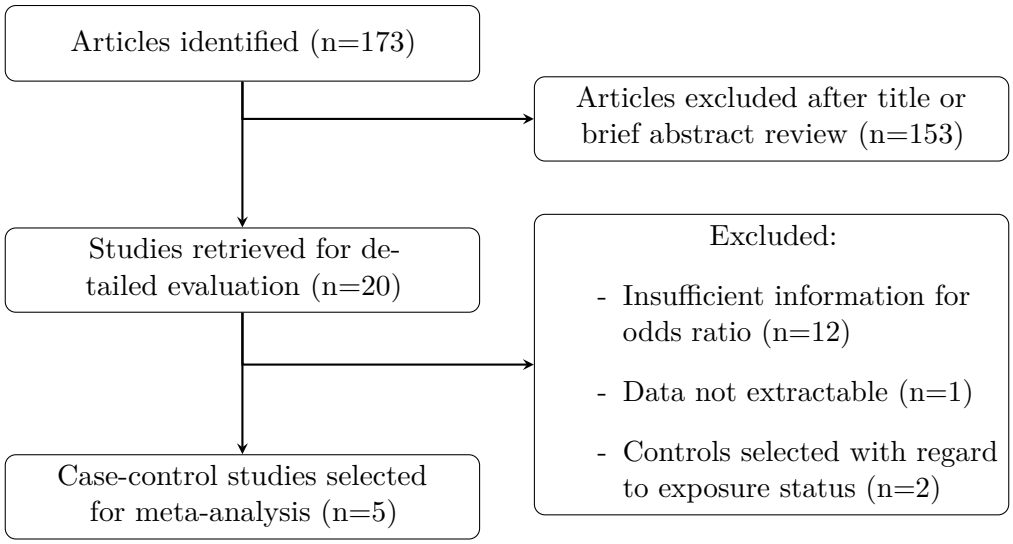

Figure 2.1: Flow chart demonstrating studies chosen for the meta-analysis.

adults whose results could not be distinguished from the children's results (studies 5-9, 11, 13-20, and 23 in Appendix B) (Figure 2.1).

2.2.3 Data Extraction

The following information was extracted from each study article: author, journal, participant ages, location, year, ethnicity, sample size, study accrual period, disease status at the time of investigation, study design, matching or adjustments, type of NS studied, exposure/type of allergy studied, primary study goal, secondary study goal, and statistics to calculate the odds ratio and corresponding confidence interval. The exposure variable included any terms used to define and characterize atopy in the studies, such as serum IgA, IgE, IgM and IgG levels, history of asthma, eczema, urticaria, hay fever, common household allergens, and allergic rhinitis. The outcome variable, nephrotic syndrome, was more uniformly defined and is consistent with KDIGO (Kidney Disease–Improving Global Outcomes) and ISKDC (International Study of Kidney Disease in Children) guidelines. Patients in selected studies were identified as having some form of idiopathic nephrotic syndrome: frequently relapsing (FRNS), steroid responsive (S.R.N.S) or minimal change disease (MCD). Steroid

dependent (SDNS), S.R.N.S., and steroid sensitive (SSNS) are sometimes used interchangeably in the literature although there are slight variations in the nature of relapse for each group (Appendix A).

2.2.4 Statistical Assessment

Because of the limited number of studies, no subgroup analysis was conducted to determine if the effect of allergy on nephrotic syndrome is consistent across the three categories of the syndrome included in this analysis or across the initial state and the relapsed state of the syndrome. We assume no distinction in characteristics between the initial state and the relapsed state regardless of the syndrome category.

Data was accumulated from sources comprising differing cultures and levels of advancement, different time periods, and different researchers operating independently. Assuming the studies are not functionally equivalent and that the effect size may differ in each study, we choose a random effects model for the analysis. The odds ratio, computed as $OR = \frac{ad}{bc}$ (Table 2.1), is used as the effect size. The within study variance in log-units is computed as $V_{Y_i} = \frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}$, where $Y_i = \ln OR$. We compute an estimate for the between-studies variance, τ^2, using the DerSimonian and Laird method (Borenstein, Hedges, Higgins, and Rothstein, 2011). This estimate is computed as

$$T^2 = \frac{Q - df}{C},$$

where $Q = \sum_{i=1}^{k} W_i Y_i^2 - \frac{\left(\sum_{i=1}^{k} W_i Y_i\right)^2}{\sum_{i=1}^{k} W_i}$, $df = k - 1$, $C = \sum_{i=1}^{k} W_i - \frac{\left(\sum_{i=1}^{k} W_i^2\right)}{\sum_{i=1}^{k} W_i}$, k is the number of studies, and W_i is the weight of the i^{th} study. The negative value of T^2 impies that the between-studies variability is 0. A Q-test for heterogeneity, formally testing the hypothesis that all studies share a common effect size, also suggests that

Table 2.1: Nomenclature of the 2 x 2 tables for nephrotic syndrome and atopy.

	Nephrotic	Non-Nephrotic	Total
Atopic	a	b	n_1
Non-atopic	c	d	n_2

the between studies variability is negligible (Q=0.59, p=0.9643, Figure 2.2). Thus, the random effects analysis is reduced to a fixed effects analysis.

The summary effect in log units is $M = \dfrac{\sum_{i=1}^{k} W_i Y_i}{\sum_{i=1}^{k} W_i}$, with the variance of the summary effect estimated as $V_M = \dfrac{1}{\sum_{i=1}^{k} W_i}$ (as the between studies variability is 0). The 95% confidence interval for the summary effect, in log units, is given by $M \pm 1.96\sqrt{V_M}$. We exponentiate the endpoints of this interval to convert to the odds ratio scale.

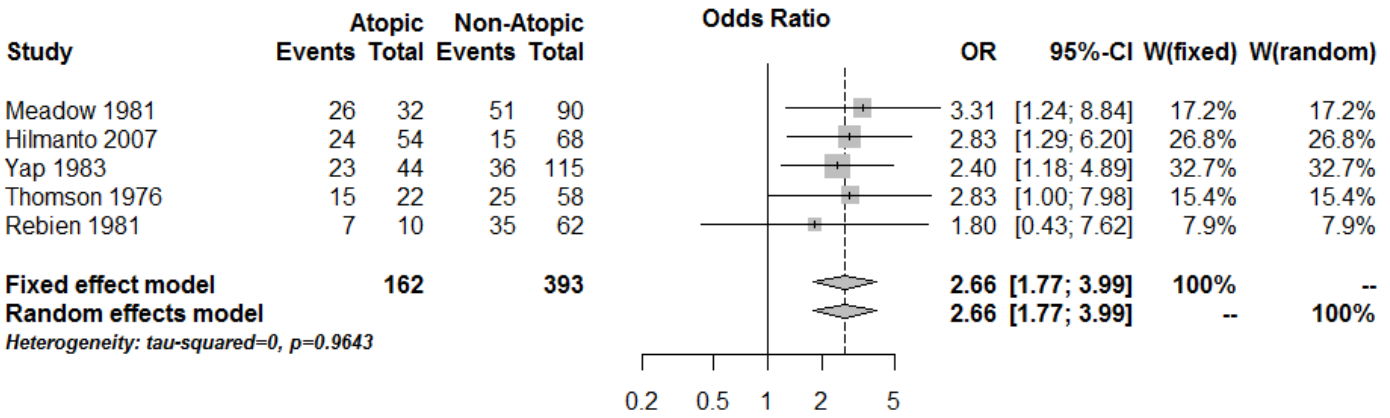

Figure 2.2: Forest plot and summary effect.

We performed all analyses using the open source software R (R Core Team, 2016). Results were considered statistically significant for two-tailed p-values < 0.05.

2.3 Results

Five case control studies were selected from the 20 full text studies comprehensively assessed. These studies were published between 1976 and 2007 with a total of 257 cases of pediatric nephrotic syndrome and 298 controls (studies 4, 10, 12, 21, and 22 in Appendix B). The included studies documented some form of atopic history and included age-matched controls (Table 2.2). In each study, information on the atopic history of the patients was obtained through parent questionnaires. Three of the five studies in the analysis used hospital or clinic-based controls (Hilmanto, 2007; Yap et al., 1983; Thomson et al., 1976), one study used healthy children from a nearby village (Meadow and Sarsfield, 1981), and one study did not indicate the source population for the controls (Rebien et al., 1981). Meta-analysis of the five case-control studies shows that a history of atopy in pediatric patients is significantly associated with higher odds of idiopathic nephrotic syndrome (OR: 2.7; 95% CI: (1.77, 3.99); Figure 2.2). The odds of NS for atopic pediatric patients is 2.7 times higher than the odds of NS for non-atopic pediatric patients.

Table 2.2: Characteristics of included studies.

First Author	Year	Location	Cases	Controls	Exposure Classification	Characteristics of Controls	OR (95% CI)	Diagnosis
Thomson	1976	London	40	40	History of asthma, eczema, or hayfever	Age-matched	2.83 (1.00,7.98)	S.R.N.S.
Meadow	1981	Leeds	77	45	History of atopy	Age, sex-matched	3.31 (1.24,8.84)	MCD S.R.N.S.
Rebien	1981	Heidelberg	42	30	History of atopy	Similar age range	1.80 (0.43,7.62)	MCD (36/42)
Yap	1983	Singapore	59	100	History of atopy	Similar age range	2.40 (1.18,4.89)	Classical S.R.N.S
Hilmanto	2007	Indonesia	39	83	History of atopy	Similar age range	2.83 (1.29,6.20)	FRNS

2.4 Discussion

The implication of an association between some form of atopy and nephrotic syndrome is well-documented (studies 4-23 in Appendix B). We used a meta-analysis to pool the results from various studies to reach a more definitive conclusion on this association. Our findings show that there is a significant association between atopy and odds of nephrotic syndrome in pediatric patients.

Other findings of the studies used in the meta-analysis are noteworthy. The Thomson study (1976) concluded that children with both Human Leukocyte Antigen (HLA-2) and a history of atopy have a risk of S.R.N.S. that is thirteen times greater than the risk for those with neither factor. The Rebien study (1981) concluded that although IgE mediated hypersensitivity in children, measured by in vitro tests, may coexist with nephrotic syndrome, it is not more prevalent than in a control population. The study also suggests that a positive atopic history should be confirmed by skin test or a radioallergosorbent test (RAST) before a subject is labeled as atopic (Rebien et al., 1981). In two later studies (Meadow and Sarsfield, 1981; Yap et al., 1983), skin tests, blood tests, RAST, and other forms of atopic identification were used contingent upon parental consent, but those results were not investigated in our analysis. The study by Meadow and Sarsfield (1981) found that some children with very high IgE levels did not have an indication of history of an atopic disorder, and no significant association existed between the frequency of certain HLAs and nephrotic syndrome. In the Hilmanto study (2007), the author concluded that HLA Class II and atopy together had an association with FRNS. These individual findings may be helpful in the continued study of pediatric nephrotic syndrome and atopy.

Patterns of relapse of the nephrotic syndrome may be significantly associated with atopy. One study reported that atopic children, particularly those suffering with eczema, relapse sooner than non-atopic children (Trompeter, Barratt, Kay, Turner, and Soothill, 1980). Another study reported that those treated at an older age

relapsed less readily (Barratt, Osofsky, Bercowsky, Soothill, and Kay, 1975), but the age effect is smaller when accounting for HLA-B12 and atopy (Trompeter et al., 1980). Further studies which quantify atopy and the severity of allergic disease may establish a predictive model for responsiveness to steroids as well as the nature of relapse. The hyper-responsive immune function may be correlated with the risk of relapse or time to initial remission, which would allow prospective stratification of individuals with nephrotic syndrome. This prospective stratification could lead to more precise treatment, which could reduce steroid toxicity and open up new domains of therapeutics.

There are several limitations to this study. To begin with, differences in the geographical locations of the studies may cause pediatric populations to vary widely. The time differences in the studies are a source of bias due to advancements in medicine and technology over the thirty-year period. In some studies, the relationship between atopic activity and nephrotic syndrome was not a primary goal, which could contribute to a form of selection bias. Particularly problematic is the assessment of the atopic activity, which varied from study to study. Furthermore, having only five studies may limit the accuracy or reliability of detecting true differences between studies (Hardy and Thompson, 1998).

The significant association detected between atopy and pediatric nephrotic syndrome warrants further study. The study of the association between atopy and pediatric nephrotic syndrome has been sporadic, and the term 'atopy' is not well-defined; however, the results presented here can lead to new ideas and hypotheses that encourage new, better defined and controlled studies. Case-control studies should be initiated with well-defined atopic parameters to further study the association between atopy and idiopathic nephrotic syndrome.

Chapter 3

Pediatric Nephrotic Syndrome Data Description

3.1 South Carolina Data

In January 2016, data was retrieved for 436 pediatric patients with over 2000 visits from the HSSC CDW database. Thirty-nine of the pediatric patients were diagnosed with pediatric nephrotic syndrome, specifically, nephrotic syndrome with unspecified pathological lesion in kidney (ICD 9 code 581.9). These diagnoses occurred between July 2007 and August 2015. There were 19 females and 20 males with ages ranging from 0-16 years at the time of diagnosis (Figure 3.1). The median age at diagnosis was five years. There were ten African-Americans, two Asians, twenty Caucasians, three Hispanic or Latino, and four classified as other or more than one race. Twenty-seven were from a medium metropolitan area, 3 from a small metropolitan area, and 9 from a non-metropolitan area. Twenty-seven of the patients were initially diagnosed in the spring and summer months (March —August), while the remaining twelve were diagnosed in the fall and winter months (September —February). Patients diagnosed between 3 and 7 years of age accounted for over half of the diagnoses.

Fourteen of the patients identified retrospectively had accompanying lab data with the date of diagnosis. One of the females is excluded from the analysis because of the limited lab data available. The 8 females and 5 males are summarized in Table 3.1.

The lab data obtained are a part of a standard comprehensive metabolic panel that can be routinely performed on patients. According to the U.S. National Library of Medicine (2016), abnormal results can be due to a variety of medical conditions,

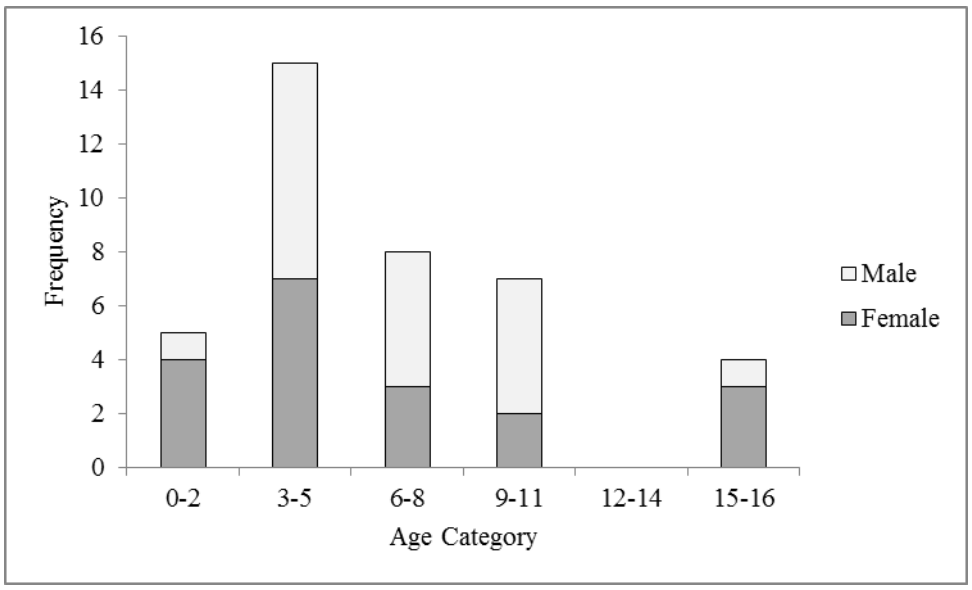

Figure 3.1: Age distribution of SC pediatric nephrotic syndrome patients at initial diagnosis.

Table 3.1: Summary lab data for 13 SC pediatric nephrotic syndrome patients.

Lab	Mean (SD)	Median	Range	UOM
Potassium Bld	5.42* (2.88)	4.6	4 - 14.7	mmol/L
Sodium Bld	138.38 (3.84)	138	133-145	mmol/L
CO2 Ser Pl	21.27 (6.89)	21	6 - 22	mmol/L
Anion	10.08 (5.11)	7	6 - 23	mmol/L
Chloride	107.92 (3.28)	107	103-113	mmol/L
BUN	31.92*(40.09)	15	8 - 129	mg/dL
Calcium	8.12 (0.86)	8.1	5.8-9.3	mg/dL
Glucose	103.85*(24.48)	94	75 - 156	mg/dL
Ser Albumin	1.78** (0.90)	1.6**	0.5 - 3.7	g/dL

* high based on normal range (U.S. National Library of Medicine, 2016)

** low based on normal range (U.S. National Library of Medicine, 2016)

including kidney failure. On average, patients in this study had elevated potassium levels, which may be indicative of kidney disorders. Patients also had elevated blood urea nitrogen (BUN) values, which could suggest the presence of kidney injury or disease, and elevated glucose levels. Levels of albumin, one of the most abundant

proteins in the body, were on average lower for the pediatric nephrotic syndrome patient. Lower serum albumin levels are indicative of the nephrotic syndrome, where the damaged kidney filtering system allows the protein to leak into the urine.

3.2 New Jersey Data

The medical records for children seen by the pediatric nephrologists at Robert Wood Johnson Medical School before March 1997 and followed for at least one year were reviewed by study authors (Constantinescu et al., 2000). There were nineteen females and thirty-four males ranging in age from 1-13 years at the time of diagnosis (Figure 3.2). Twenty-five of the patients received a diagnosis of SDNS, nine were diagnosed with FRNS, seventeen with infrequently relapsing nephrotic syndrome (IFRNS) and two with SRNS. The median age at diagnosis was 3.5 years. The median number of days to remission was 10, with remission defined by the study as protein-free urine. The initial study reported a race distribution of 76.9% white, 8.9% black, 7.1% Hispanic and 7.1% other.

Reported lab data included cholesterol level, creatinine level and the presence or absence of hematuria. A summary of the lab data is presented in Table 3.2. Thirty of the patients showed no hematuria at initial diagnosis, sixteen had micro-hematuria, and seven exhibited macro-hematuria. The average cholesterol level for the group was much higher than the upper bound of the normal range for cholesterol levels in children and adolescents (American Academy of Pediatrics, 2015). High cholesterol may result from a number of conditions, including kidney disease (Dietz and Stern, 2011).

3.3 Discussion

Pediatric nephrotic syndrome is classified as a part of the Rare Disease Clinical Research Network (National Institute of Health, 2016). The pediatric nephrotic

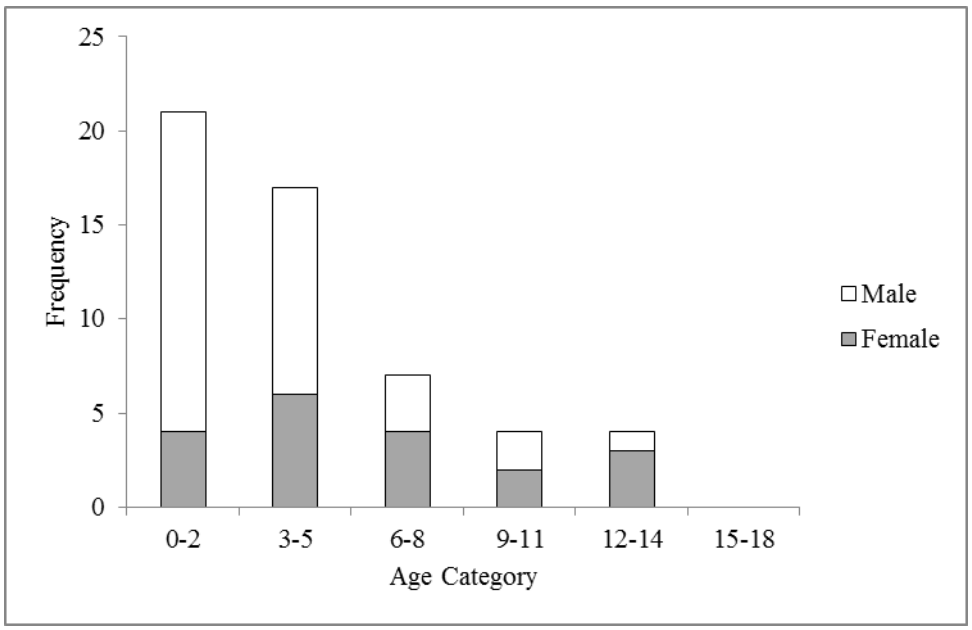

Figure 3.2: Age distribution of NJ pediatric nephrotic syndrome patients at initial diagnosis.

Table 3.2: Summary lab data for NJ pediatric nephrotic syndrome patients.

Lab	Mean (SD)	Median	Range	UOM
Cholesterol	424.16* (18.85)	387	201-799	mg/dL
Creatinine	0.49 (0.03)	0.5	0.1-1.0	mg/dL

* high based on normal range for children (American Academy of Pediatrics, 2015)

syndrome data studied here are the result of retrospective chart reviews and are limited by access to clinical data that have been recorded and are available.

Chapter 4

The reflected-shifted-truncated-gamma distribution with application to negatively skewed survival data

4.1 Introduction

The two-parameter gamma distribution has been used extensively in survival analysis. It is useful for modeling survival processes that do not fit into a symmetric distribution (X. Liu, 2012). Although flexibility is allowed by this unimodal two-parameter distribution, the basic shape ranges from positively skewed for small values of the shape parameter to approximately normal for large values of the shape parameter for a fixed value of the scale parameter (Johnson, Kotz, and Balakrishnan, 2002; Ofungwu, 2014). The Weibull, exponential, lognormal, and normal distributions are alternative standard distributions commonly employed to model data that is approximately normal to positively skewed (Hougaard, 1999).

The generalized gamma distribution is a three-parameter distribution that was first presented by Stacy (1962) and includes as special sub-models the exponential, Weibull, gamma and Rayleigh distributions. Variations of the generalized gamma distribution have been proposed in recent years to enhance its modeling capability. These include the Kumaraswamy generalized gamma distribution (de Pascoa, Ortega, and Cordeiro, 2011), the exponentiated generalized gamma distribution (Cordeiro, Ortega, and Silva, 2011), and the transmuted generalized gamma distribution (Lucena,

Silva, and Cordeiro, 2015). These variations were designed to provide more flexibility to the gamma distribution by allowing the capability of modeling both monotone and non-monotone failure rates (Lucena et al., 2015). Despite the improved flexibility, the distributions are generalizations of the standard two-parameter gamma distribution and are still mainly utilized for positively skewed data.

The Gompertz distribution is a standard distribution for modeling negatively skewed survival data. It was originally developed in 1825 to model human mortality (Gompertz, 1825). A major drawback of the Gompertz distribution is that it fits only adult mortality sufficiently (Thatcher, 1999). Several variations or extensions of the Gompertz distribution have also been introduced in response to the modeling of human mortality data (Cooray and Ananda, 2010).

Aside from the Gompertz distribution and its extensions, variations of the normal distribution have been proposed to model negatively skewed data. These include the skew normal (Azzalini, 1985), the power normal (Gupta and Gupta, 2008), the tilted normal (Maiti and Dey, 2012), and a generalized normal distribution (Robertson and Allison, 2012). Nevertheless, the applicability of these distributions is limited. Practical difficulties of estimating the skewness parameter for small to moderate sample sizes have been noted with the skew normal distribution, as well as problems with goodness of fit for the power normal distribution (Maiti and Dey, 2012). The tilted normal distribution is derived using a Marshall-Olkin transformation to induce skewness; however, this transformation applied to a unimodal symmetric density results in a distribution that is not flexible enough to handle data presenting high or moderate skewness (Rubio and Steel, 2012). The generalized normal distribution was constructed to model human longevity and distributional properties involve constraints relevant only to life table data (Robertson and Allison, 2012).

We propose a reflected, shifted, truncated version of the two-parameter gamma distribution as an alternative distribution for modeling negatively skewed survival data

and demonstrate the applicability of this distribution using three types of survival data: complete, where the event of interest is observed exactly; right-censored, where the event of interest is only known to have not occurred by a given time point; and interval-censored, where the event of interest is only known to have occurred in a particular interval of time.

4.2 The Reflected-Shifted-Truncated-Gamma Distribution

The reflected-shifted-truncated-gamma (RSTG) distribution is constructed through a series of transformations to the two-parameter gamma distribution. The probability density function of the two-parameter gamma distribution is

$$f(t|\alpha,\theta) = \frac{1}{\Gamma(\alpha)\theta^{\alpha}} e^{-\frac{t}{\theta}} t^{\alpha-1}, 0 < t < \infty.$$

Here, $\alpha > 0$ represents the shape parameter and $\theta > 0$ represents the scale parameter. Reflecting the two-parameter gamma distribution about the y-axis and shifting it $k > 0$ units to the right gives a probability density function of

$$f_1(t|\alpha,\theta,k) = \frac{1}{\Gamma(\alpha)\theta^{\alpha}} e^{-\frac{(-t+k)}{\theta}} (-t+k)^{\alpha-1}, -\infty < t < k; \alpha,\theta > 0.$$

The cumulative distribution function of this three-parameter reflected, shifted gamma distribution is

$$\begin{aligned} F_1(t|\alpha,\theta,k) &= \int_{-\infty}^{t} \frac{1}{\Gamma(\alpha)\theta^{\alpha}} e^{-\frac{(-x+k)}{\theta}} (-x+k)^{\alpha-1} \mathrm{d}x \\ &= \frac{1}{\Gamma(\alpha)} \left[\Gamma\left(\alpha, \frac{-t+k}{\theta}\right) \right] \end{aligned}$$

for $t < k$ where $\Gamma(a) = \int_0^{\infty} t^{a-1} e^{-t}\,\mathrm{d}t$ is the gamma function and $\Gamma(a,b) = \int_b^{\infty} t^{a-1} e^{-t}\,\mathrm{d}t$ represents the upper incomplete gamma function.

Truncating the reflected, shifted gamma distribution at 0 effectively restricts the new distribution to the interval $[0, k)$. The probability density function for this RSTG

distribution is

$$f^*(t|\alpha,\theta,k) = \frac{1}{F_1(k) - F_1(0)} \left(\frac{1}{\Gamma(\alpha)\theta^\alpha}\right) e^{-\frac{(-t+k)}{\theta}} (-t+k)^{\alpha-1}$$
$$= \frac{1}{\theta^\alpha \left(\Gamma(\alpha) - \Gamma(\alpha, \frac{k}{\theta})\right)} e^{-\frac{(-t+k)}{\theta}} (-t+k)^{\alpha-1} \quad \alpha > 0,\ \theta > 0,\ 0 \le t < k, \tag{4.1}$$

with cumulative distribution function given by

$$F^*(t|\alpha,\theta,k) = \frac{1}{\theta^\alpha \left(\Gamma(\alpha) - \Gamma(\alpha, \frac{k}{\theta})\right)} \int_0^t e^{-\frac{(-x+k)}{\theta}} (-x+k)^{\alpha-1}\, \mathrm{d}x$$
$$= \frac{\Gamma\left(\alpha, \frac{-t+k}{\theta}\right) - \Gamma\left(\alpha, \frac{k}{\theta}\right)}{\Gamma(\alpha) - \Gamma\left(\alpha, \frac{k}{\theta}\right)}.$$

Complete derivations of the density and distribution functions are given in Appendix C.

Plots of the probability density function for values of α, θ and k are given in Figure 4.1.

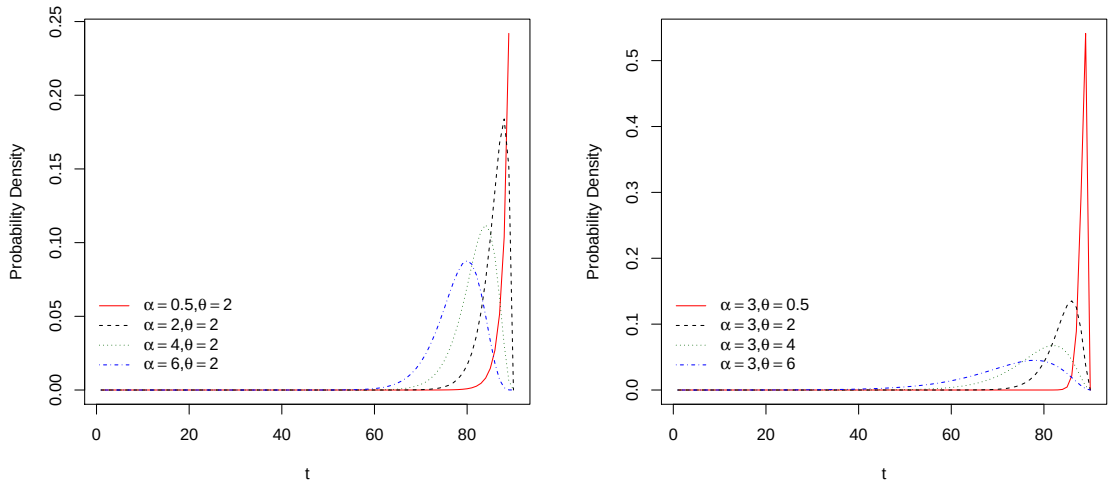

Figure 4.1: Probability density function of the RSTG distribution with varying α, θ; k=90.

4.2.1 Quantile Function and Moments of the RSTG Distribution

The quantile function of the RSTG distribution is defined as $Q(p) = \inf\{t : F^*(t) \geq p\}$ for $p \in (0, 1]$. It can be obtained by solving the following equation for t:

$$\Gamma\left(\alpha, \frac{-t+k}{\theta}\right) = p\left[\Gamma(\alpha) - \Gamma\left(\alpha, \frac{k}{\theta}\right)\right] + \Gamma\left(\alpha, \frac{k}{\theta}\right). \tag{4.2}$$

For a given $p \in (0, 1]$, $t_p = Q(p)$ represents the $100p^{th}$ percentile.

The quantile function of the distribution can be used to construct quantile analogs of standard moment-based descriptive measures and to extend those standard descriptive measures (Gilchrist, 2000). We can use this function to generate random data that describe the density given in Equation (4.1).

The n^{th} raw moment of the RSTG distribution is given by

$$\begin{aligned} E[T^n] &= \frac{1}{\theta^\alpha\left(\Gamma(\alpha) - \Gamma(\alpha, \frac{k}{\theta})\right)} \int_0^k t^n e^{-\frac{(-t+k)}{\theta}} (-t+k)^{\alpha-1}\,\mathrm{d}t \\ &= \sum_{j=0}^{n} (-1)^j \binom{n}{n-j} k^{n-j}\theta^j \frac{\left(\Gamma\left(\alpha+j\right) - \Gamma\left(\alpha+j, \frac{k}{\theta}\right)\right)}{\Gamma(\alpha) - \Gamma(\alpha, \frac{k}{\theta})}. \end{aligned}$$

In particular, the first moment is

$$\begin{aligned} E[T] &= \frac{1}{\theta^\alpha\left(\Gamma(\alpha) - \Gamma(\alpha, \frac{k}{\theta})\right)} \int_0^k t e^{-\frac{(-t+k)}{\theta}} (-t+k)^{\alpha-1}\,\mathrm{d}t \\ &= k - \frac{\theta\left(\Gamma(\alpha+1) - \Gamma\left(\alpha+1, \frac{k}{\theta}\right)\right)}{\left(\Gamma(\alpha) - \Gamma\left(\alpha, \frac{k}{\theta}\right)\right)} \end{aligned}$$

and the second moment is

$$\begin{aligned} E[T^2] &= \frac{1}{\theta^\alpha\left(\Gamma(\alpha) - \Gamma(\alpha, \frac{k}{\theta})\right)} \int_0^k t^2 e^{-\frac{(-t+k)}{\theta}} (-t+k)^{\alpha-1}\,\mathrm{d}t \\ &= k^2 - 2k\theta\left[\frac{\theta\left(\Gamma(\alpha+1) - \Gamma\left(\alpha+1, \frac{k}{\theta}\right)\right)}{\left(\Gamma(\alpha) - \Gamma\left(\alpha, \frac{k}{\theta}\right)\right)}\right] + \theta^2\left[\frac{\theta\left(\Gamma(\alpha+2) - \Gamma\left(\alpha+2, \frac{k}{\theta}\right)\right)}{\left(\Gamma(\alpha) - \Gamma\left(\alpha, \frac{k}{\theta}\right)\right)}\right] \end{aligned}$$

where $\alpha > 0$, $\theta > 0$, and $k > 0$. Hence, the mean of the RSTG distribution is

$$E(T) = k - \frac{\theta\left(\Gamma(\alpha+1) - \Gamma\left(\alpha+1, \frac{k}{\theta}\right)\right)}{\left(\Gamma(\alpha) - \Gamma\left(\alpha, \frac{k}{\theta}\right)\right)},$$

with variance

$$V(T) = E[T^2] - (E[T])^2$$
$$= \theta^2 \left[\left(\frac{\Gamma(\alpha+2) - \Gamma\left(\alpha+2, \frac{k}{\theta}\right)}{\Gamma(\alpha) - \Gamma\left(\alpha, \frac{k}{\theta}\right)} \right) - \left(\frac{\Gamma(\alpha+1) - \Gamma\left(\alpha+1, \frac{k}{\theta}\right)}{\Gamma(\alpha) - \Gamma\left(\alpha, \frac{k}{\theta}\right)} \right)^2 \right].$$

Pearson's coefficient of skewness (CS) and kurtosis (CK) are given by

$$CS = E\left[\left(\frac{T-\mu}{\sigma}\right)^3\right]$$
$$= \frac{E[T^3] - 3\mu E[T^2] + 2\mu^3}{\sigma^3}$$

$$CK = E\left[\left(\frac{T-\mu}{\sigma}\right)^4\right]$$
$$= \frac{E[T^4] - 4\mu E[T^3] + 6\mu^2 E[T^2] - 3\mu^4}{\sigma^4}$$

where $\mu = E[T]$. Plots of the mean, variance, skewness, and kurtosis functions of the RSTG distribution with $\alpha = 1$ and $k = 96$ are shown in Figure 4.2.

4.2.2 Survival and Hazard Functions of the RSTG Distribution

The survival and hazard functions of the RSTG distribution are

$$S^*(t) = 1 - F^*(t) = \frac{\Gamma(\alpha) - \Gamma\left(\alpha, \frac{-t+k}{\theta}\right)}{\Gamma(\alpha) - \Gamma\left(\alpha, \frac{k}{\theta}\right)}, \quad \text{for } \alpha > 0,\ \theta > 0,\ 0 < t < k \tag{4.3}$$

and

$$h^*(t) = \frac{f^*(t)}{S^*(t)} = \frac{e^{-\frac{(-t+k)}{\theta}}(-t+k)^{\alpha-1}}{\theta^\alpha \left(\Gamma(\alpha) - \Gamma\left(\alpha, \frac{-t+k}{\theta}\right)\right)}, \quad \text{for } \alpha > 0,\ \theta > 0,\ 0 < t < k \tag{4.4}$$

respectively. Plots of the survival and hazard functions for values of α and θ for a fixed k are shown in Figure 4.3.

4.3 Parametric Estimation

Both frequentist and Bayesian approaches can be used to estimate the parameters of a survival model (Pradhan and Kundu, 2011). Typically, parametric estimation follows

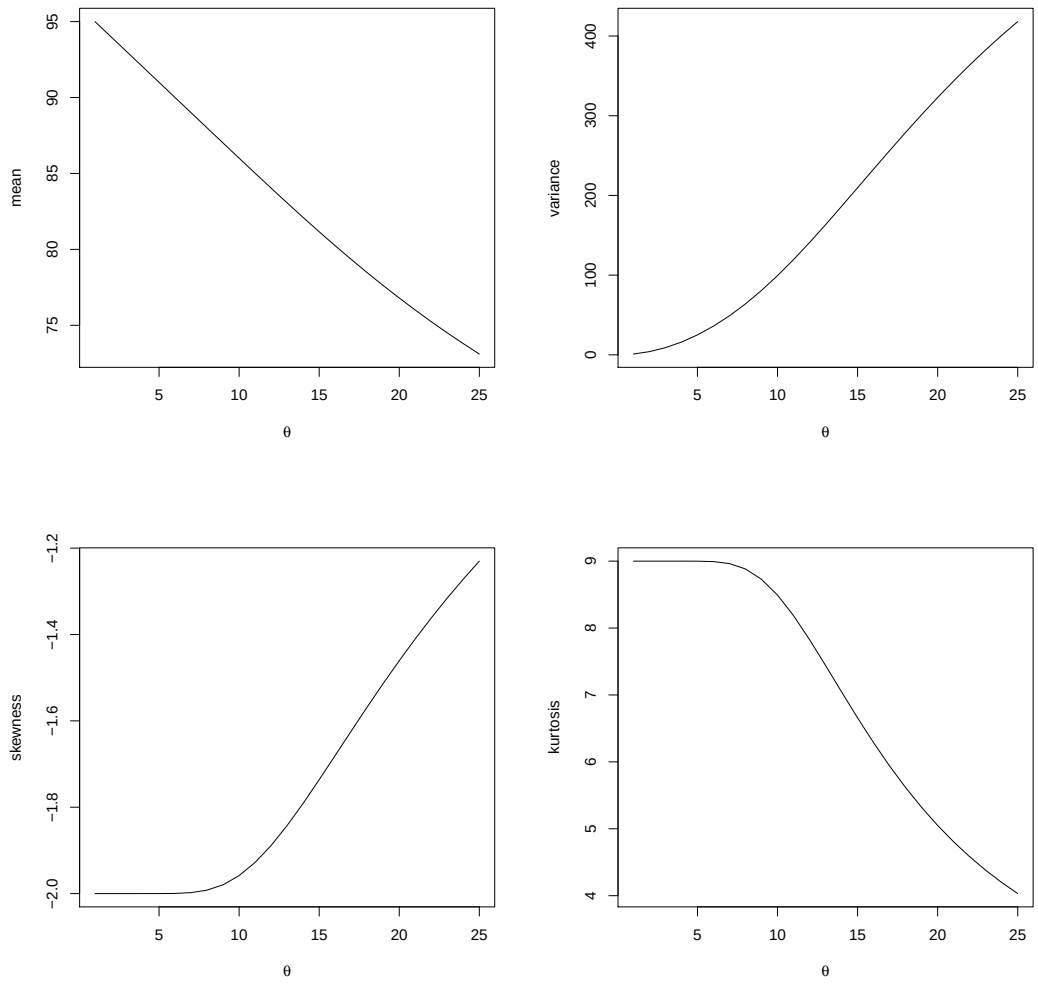

Figure 4.2: The mean, variance, skewness and kurtosis of the RSTG distribution with $\alpha = 1$ and $k = 90$.

the frequentist approach and is based on likelihood methods (Lee and Wang, 2003; Kalbfleisch and Prentice, 2011). Parametric estimation for the RSTG distribution will use the method of maximizing the log-likelihood function.

4.3.1 Maximum Likelihood Estimation for Complete Data

Let $t_1, t_2, ..., t_n$ be a random sample of size n with probability density function given by equation (4.1). The likelihood function for the parameter vector $\Theta = (\alpha, \theta, k)$ based on the observed sample is proportional to

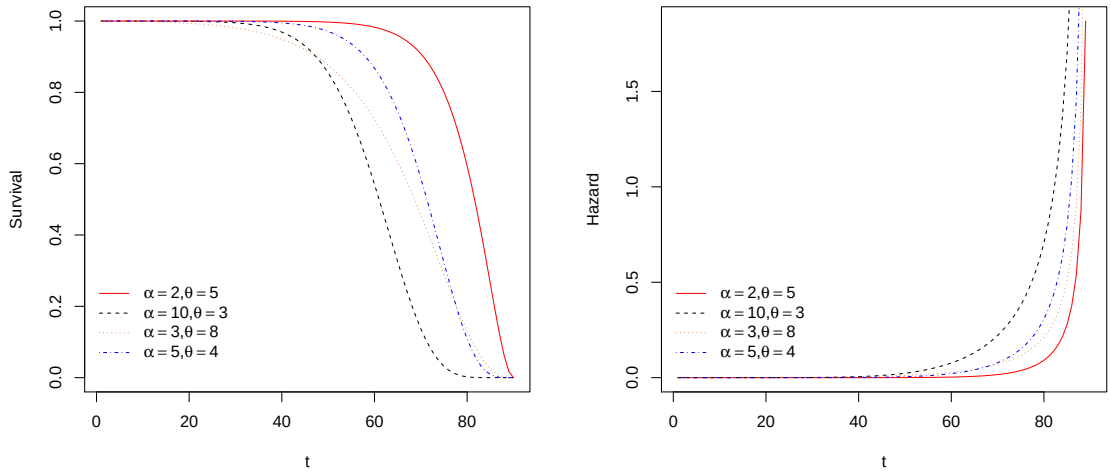

Figure 4.3: Survival function and hazard function of the RSTG distribution with various α, θ; $k = 90$.

$$
\begin{aligned}
L(\Theta) &= \prod_{i=1}^{n} f^*(t_i|\Theta) \\
&= \frac{1}{\left(\theta^{\alpha}\left(\Gamma(\alpha) - \Gamma(\alpha, \frac{k}{\theta})\right)\right)^n} e^{-\frac{\sum_{i=1}^{n}(-t_i+k)}{\theta}} \prod_{i=1}^{n}(-t_i+k)^{\alpha-1}.
\end{aligned}
$$

Without loss of generality, the corresponding log-likelihood function is written as

$$
\begin{aligned}
l(\Theta) &= \ln L(\Theta) \\
&= -n\left[\alpha \ln\theta + \ln\left(\Gamma(\alpha) - \Gamma\left(\alpha, \frac{k}{\theta}\right)\right)\right] - \frac{\sum_{i=1}^{n}(-t_i+k)}{\theta} \\
&+ (\alpha-1)\sum_{i=1}^{n}\ln(-t_i+k).
\end{aligned} \tag{4.5}
$$

We assume that the parameters α, θ and k are unknown. We obtain the normal equations for the unknown parameters by taking partial derivatives of equation (4.5) with respect to α, θ and k and equating each to zero. The resulting equations, in which $l = l(\Theta)$, are

$$
\frac{dl}{d\alpha} = -n\left\{\ln\theta + \frac{\left[\Gamma'(\alpha) - \Gamma'\left(\alpha, \frac{k}{\theta}\right)\right]}{\left[\Gamma(\alpha) - \Gamma\left(\alpha, \frac{k}{\theta}\right)\right]}\right\} + \sum_{i=1}^{n}\ln(-t_i+k) = 0, \tag{4.6}
$$

where $\Gamma'(\alpha) = \frac{d\Gamma(\alpha)}{d\alpha} = \frac{d}{d\alpha}(\ln\Gamma(\alpha))\Gamma(\alpha)$, and $\Gamma'(\alpha, \frac{k}{\theta}) = \frac{d\Gamma(\alpha,\frac{k}{\theta})}{d\alpha} = \int_{\frac{k}{\theta}}^{\infty} \ln(y) y^{\alpha-1} e^{-y} \mathrm{d}y$,

$$\frac{dl}{d\theta} = -n\left\{\frac{\alpha}{\theta} + \frac{-\frac{1}{\theta}\left(\frac{k}{\theta}\right)^{\alpha} e^{-\frac{k}{\theta}}}{\Gamma(\alpha) - \Gamma\left(\alpha, \frac{k}{\theta}\right)}\right\} + \frac{\sum_{i=1}^{n}(-t_i + k)}{\theta^2} = 0. \tag{4.7}$$

and

$$\frac{dl}{dk} = \frac{-\frac{n}{k}\left(\frac{k}{\theta}\right)^{\alpha} e^{-\frac{k}{\theta}}}{\Gamma(\alpha) - \Gamma\left(\alpha, \frac{k}{\theta}\right)} - \frac{n}{\theta} + \sum_{i=1}^{n} \frac{\alpha - 1}{(-t_i + k)} = 0. \tag{4.8}$$

The solutions of equations (4.6), (4.7) and (4.8) are candidates for the maximum likelihood estimates (MLEs) of the parameters α, θ and k. The maximum is attained at the candidate values $\hat{\alpha}, \hat{\theta}$, and $\hat{k}$ if the Hessian matrix (Appendix D) is negative definite at those candidate values.

Although the MLEs of the unknown parameters can be obtained, we cannot obtain the exact distribution of the MLEs. We use the large sample approximation. Assuming regularity conditions are satisfied, the asymptotic confidence intervals can be obtained by using the observed Fisher information matrix.

For parameter vector $\Theta = (\alpha, \theta, k)$, the observed Fisher information matrix is given by

$$I(\Theta) = -\begin{pmatrix} H_{11}(\Theta) & H_{12}(\Theta) & H_{13}(\Theta) \\ H_{21}(\Theta) & H_{22}(\Theta) & H_{23}(\Theta) \\ H_{31}(\Theta) & H_{32}(\Theta) & H_{33}(\Theta) \end{pmatrix} = \begin{pmatrix} a & b & c \\ b & d & e \\ c & e & f \end{pmatrix},$$

where $H_{ij}(\Theta)$ represents the ij^{th} entry of the Hessian matrix (Appendix D). The variance-covariance matrix of the parameter estimates can be approximated by the inverse of the information matrix

$$I^{-1}(\theta) = \frac{1}{K} \cdot \begin{pmatrix} a & b & c \\ b & d & e \\ c & e & f \end{pmatrix}^{*}.$$

where $K = a(df - e^2) - b(bf - ec) + c(be - cd)$ and $()^*$ represents the adjoint of the matrix. The standard errors of the estimates are given by

$$\sigma_{\hat{\alpha}} = \sqrt{\frac{df - e^2}{K}}, \quad \sigma_{\hat{\theta}} = \sqrt{\frac{af - c^2}{K}}, \quad \text{and} \quad \sigma_{\hat{k}} = \sqrt{\frac{ad - b^2}{K}}. \tag{4.9}$$

We can use equation (4.9) to derive approximate $100(1-\tau)\%$ confidence intervals for the MLEs of the parameters α, θ and k of the forms

$$\left(\hat{\alpha} \pm z_{\frac{\tau}{2}}\sigma_{\hat{\alpha}}\right), \left(\hat{\theta} \pm z_{\frac{\tau}{2}}\sigma_{\hat{\theta}}\right), \text{and } \left(\hat{k} \pm z_{\frac{\tau}{2}}\sigma_{\hat{\theta}}\right) \tag{4.10}$$

where $z_{\tau/2}$ is the upper $100(\tau/2)^{th}$ percentile of the standard normal distribution.

The MLEs cannot be solved for explicitly here and must be found by numeric methods such as Newton-Raphson's algorithm. Details of the Newton-Raphson algorithm and other numeric methods can be found in textbooks on numerical methods for optimization, such as *Iterative Methods for Optimization* by Carl Kelley (1999). We use iterative methods in **R** to obtain the MLEs and standard errors.

4.3.2 Maximum Likelihood Estimation for Right-Censored Data

Let $t_1, t_2, ..., t_n$ be a right-censored random sample of size n with probability density function given by equation (4.1). The censoring indicator δ_i is such that

$$\delta_i = \begin{cases} 1 & \text{if the event of interest is observed} \\ 0 & \text{if the event of interest is not observed (event time is right-censored)} \end{cases}.$$

The likelihood function for right-censored data is proportional to

$$\begin{aligned} L(\Theta) &= \prod_{i=1}^{n} f^*(t_i|\Theta)^{\delta_i} S^*(t_i|\Theta)^{1-\delta_i} \\ &= \prod_{i=1}^{n} \left[\frac{1}{\theta^\alpha \left(\Gamma(\alpha) - \Gamma(\alpha, \frac{k}{\theta})\right)} e^{-\frac{(-t_i+k)}{\theta}} (-t_i + k)^{\alpha-1} \right]^{\delta_i} \left[\frac{\Gamma(\alpha) - \Gamma\left(\alpha, \frac{-t_i+k}{\theta}\right)}{\Gamma(\alpha) - \Gamma\left(\alpha, \frac{k}{\theta}\right)} \right]^{1-\delta_i} \end{aligned}$$

and the corresponding log-likelihood function is

$$
\begin{aligned}
l(\Theta) &= \ln L(\Theta) \\
&= \sum_{i=1}^{n} \delta_i \Bigg\{ - \left[\alpha \ln \theta + \ln \left(\Gamma(\alpha) - \Gamma\left(\alpha, \frac{k}{\theta}\right)\right)\right] - \frac{-t_i + k}{\theta} + (\alpha - 1)\ln(-t_i + k) \Bigg\} \\
&\quad + \sum_{i=1}^{n} (1 - \delta_i) \left\{ \ln \left[\Gamma(\alpha) - \Gamma\left(\alpha, \frac{-t_i + k}{\theta}\right)\right] - \ln \left[\Gamma(\alpha) - \Gamma\left(\alpha, \frac{k}{\theta}\right)\right] \right\}.
\end{aligned} \tag{4.11}
$$

We calculate partial derivatives of equation (4.11) with respect to α, θ and k to obtain the normal equations for the unknown parameters and equate each to zero. The resulting equations are

$$
\begin{aligned}
\frac{dl}{d\alpha} = \sum_{i=1}^{n} \delta_i &\left(- \ln \theta - \frac{\Gamma'(\alpha) - \Gamma'\left(\alpha, \frac{k}{\theta}\right)}{\Gamma(\alpha) - \Gamma\left(\alpha, \frac{k}{\theta}\right)} + \ln(-t_i + k) \right) \\
&+ \sum_{i=1}^{n} (1 - \delta_i) \left(\frac{\Gamma'(\alpha) - \Gamma'\left(\alpha, \frac{-t_i+k}{\theta}\right)}{\Gamma(\alpha) - \Gamma\left(\alpha, \frac{-t_i+k}{\theta}\right)} - \frac{\Gamma'(\alpha) - \Gamma'\left(\alpha, \frac{k}{\theta}\right)}{\Gamma(\alpha) - \Gamma\left(\alpha, \frac{k}{\theta}\right)} \right) = 0,
\end{aligned}
$$

with $\Gamma'(\alpha)$, and $\Gamma'(\alpha, \frac{k}{\theta})$ as defined in equation (4.6),

$$
\begin{aligned}
\frac{dl}{d\theta} = \sum_{i-1}^{n} \delta_i &\left(-\frac{\alpha}{\theta} + \frac{\frac{1}{\theta}\left(\frac{k}{\theta}\right)^{\alpha} e^{-\frac{k}{\theta}}}{\Gamma(\alpha) - \Gamma\left(\alpha, \frac{k}{\theta}\right)} + \frac{-t_i + k}{\theta^2} \right) \\
&+ \sum_{i-1}^{n} (1 - \delta_i) \left(\frac{-\frac{1}{\theta}\left(\frac{-t_i+k}{\theta}\right)^{\alpha} e^{-\frac{-t_i+k}{\theta}}}{\Gamma(\alpha) - \Gamma\left(\alpha, \frac{-t_i+k}{\theta}\right)} + \frac{\frac{1}{\theta}\left(\frac{k}{\theta}\right)^{\alpha} e^{-\frac{k}{\theta}}}{\Gamma(\alpha) - \Gamma\left(\alpha, \frac{k}{\theta}\right)} \right) = 0,
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{dl}{dk} = \sum_{i=1}^{n} \delta_i &\left(\frac{-\frac{1}{k}\left(\frac{k}{\theta}\right)^{\alpha} e^{-\frac{k}{\theta}}}{\Gamma(\alpha) - \Gamma\left(\alpha, \frac{k}{\theta}\right)} - \frac{1}{\theta} + \frac{(\alpha - 1)}{-t_i + k} \right) \\
&+ \sum_{i=1}^{n} (1 - \delta_i) \left(\frac{\frac{1}{-t_i+k}\left(\frac{-t_i+k}{\theta}\right)^{\alpha} e^{-\frac{-t_i+k}{\theta}}}{\Gamma(\alpha) - \Gamma\left(\alpha, \frac{-t_i+k}{\theta}\right)} - \frac{\frac{1}{k}\left(\frac{k}{\theta}\right)^{\alpha} e^{-\frac{k}{\theta}}}{\Gamma(\alpha) - \Gamma\left(\alpha, \frac{k}{\theta}\right)} \right) = 0.
\end{aligned}
$$

We obtain the MLEs, standard errors and confidence intervals using methods discussed in Section 4.3.1.

4.3.3 Maximum Likelihood Estimation for Interval-Censored Data

Let $t_1, t_2, ..., t_n$ be a random sample such that $t_i \in [l_i, r_i)$, $l_i \leq r_i$, where l_i is the left limit of the i^{th} censored data point and r_i is the right limit of the i^{th} censored data point. The likelihood function for interval-censored data is given by

$$\begin{aligned} L(\Theta) &= \prod_{i=1}^{n} \left(S^*(l_i|\Theta) - S^*(r_i|\Theta)\right) \\ &= \frac{1}{\left(\Gamma(\alpha) - \Gamma\left(\alpha, \frac{k}{\theta}\right)\right)^n} \prod_{i=1}^{n} \left(\Gamma\left(\alpha, \frac{-r_i + k}{\theta}\right) - \Gamma\left(\alpha, \frac{-l_i + k}{\theta}\right)\right) \end{aligned}$$

The log-likelihood can be written as

$$\begin{aligned} l(\Theta) &= \ln L(\Theta) \\ &= -n \ln\left(\Gamma(\alpha) - \Gamma\left(\alpha, \frac{k}{\theta}\right)\right) + \sum_{i=1}^{n} \ln\left(\Gamma\left(\alpha, \frac{-r_i + k}{\theta}\right) - \Gamma\left(\alpha, \frac{-l_i + k}{\theta}\right)\right) \end{aligned} \tag{4.12}$$

To obtain the normal equations for the unknown parameters, we calculate partial derivatives of equation (4.12) with respect to α, θ and k and equate each to zero. The resulting equations are

$$\frac{dl}{d\alpha} = -n\left(\frac{\Gamma'(\alpha) - \Gamma'\left(\alpha, \frac{k}{\theta}\right)}{\Gamma(\alpha) - \Gamma\left(\alpha, \frac{k}{\theta}\right)}\right) + \sum_{i=1}^{n} \frac{\Gamma'\left(\alpha, \frac{-r_i+k}{\theta}\right) - \Gamma'\left(\alpha, \frac{-l_i+k}{\theta}\right)}{\Gamma\left(\alpha, \frac{-r_i+k}{\theta}\right) - \Gamma\left(\alpha, \frac{-l_i+k}{\theta}\right)} = 0,$$

with $\Gamma'(\alpha)$, and $\Gamma'(\alpha, \frac{k}{\theta})$ as defined in equation (4.6),

$$\frac{dl}{d\theta} = -n\left(\frac{-\frac{1}{\theta}\left(\frac{k}{\theta}\right)^{\alpha} e^{-\frac{k}{\theta}}}{\Gamma(\alpha) - \Gamma\left(\alpha, \frac{k}{\theta}\right)}\right) + \sum_{i=1}^{n} \left(\frac{\frac{1}{\theta}\left(\frac{-r_i+k}{\theta}\right)^{\alpha} e^{-\frac{-r_i+k}{\theta}} - \frac{1}{\theta}\left(\frac{-l_i+k}{\theta}\right)^{\alpha} e^{-\frac{-l_i+k}{\theta}}}{\Gamma(\alpha, \frac{-r_i+k}{\theta}) - \Gamma\left(\alpha, \frac{-l_i+k}{\theta}\right)}\right) = 0,$$

and

$$\frac{dl}{dk} = \frac{-\frac{n}{k}\left(\frac{k}{\theta}\right)^{\alpha} e^{-\frac{k}{\theta}}}{\Gamma(\alpha) - \Gamma\left(\alpha, \frac{k}{\theta}\right)} + \sum_{i=1}^{n} \left(\frac{\frac{-1}{-r_i+k}\left(\frac{-r_i+k}{\theta}\right)^{\alpha} e^{-\frac{-r_i+k}{\theta}} + \frac{1}{-l_i+k}\left(\frac{-l_i+k}{\theta}\right)^{\alpha} e^{-\frac{-l_i+k}{\theta}}}{\Gamma(\alpha, \frac{-r_i+k}{\theta}) - \Gamma\left(\alpha, \frac{-l_i+k}{\theta}\right)}\right) = 0.$$

We again obtain the MLEs, standard errors and confidence intervals using methods discussed in Section 4.3.1.

4.4 Simulations

A series of Monte Carlo simulations are conducted to assess the performance and consistency of the maximum likelihood estimators for the RSTG distribution. Bias and mean squared error (MSE) criteria are used for comparison purposes. For each of 1000 samples, we generate $n = 20, 40, 60, 80, 100$ and 200 random variables $T_i, i = 1...n$, from the RSTG distribution with shape parameter $\alpha = 2$, scale parameter $\theta = 2$, and shift parameter $k = 96$ using equation (4.2).

To generate right-censored data values, administrative censoring is used following a method by Michael and Lambert (Crowther and Lambert, 2013) so that the censoring percentage is approximately $10 - 20\%$. To generate interval-censored data, we generate n values from the RSTG distribution to serve as left endpoints of each interval. We sort the values and use a pre-specified probability from the uniform distribution to determine which of the two adjacent ordered values will serve as the right endpoint. We fit each complete, right-censored, and interval-censored sample using the RSTG distribution.

The mean values of the parameter estimates, bias, and MSE for each sample size are presented in Table 4.1. Standard errors were calculated using the bootstrap method.

Table 4.1: Estimated standard error, bias and MSE of MLE of parameters based on 1000 simulations of complete, right-censored and interval-censored data of the RSTG(2,2,96) distribution with n=20, 40, 60, 80, 100, and 200.

		Complete			Right-censored			Interval-censored		
n		MLE (se)	Bias	MSE	MLE (se)	Bias	MSE	MLE (se)	Bias	MSE
20	α	1.63 (1.59)	-0.37	2.653	3.41 (1.74)	1.41	4.990	3.78 (2.44)	1.78	9.118
	θ	2.94 (1.23)	0.94	2.410	2.20 (1.16)	0.20	1.387	1.61 (0.81)	-0.39	0.808
	k	95.65 (0.62)	-0.35	0.504	96.63 (1.15)	0.63	1.729	95.62 (0.65)	-0.38	0.560
40	α	1.87 (1.49)	-0.13	2.233	3.04 (1.73)	1.03	4.990	2.99 (2.14)	0.99	5.562
	θ	2.49 (1.28)	0.49	1.881	1.94 (0.82)	-0.07	0.679	1.77 (0.74)	-0.23	0.607
	k	95.85 (0.59)	- 0.15	0.372	96.49 (0.95)	0.49	1.145	95.67 (0.62)	-0.33	0.488
60	α	1.89 (0.93)	-0.11	0.885	2.66 (1.63)	0.66	3.109	2.65 (1.90)	0.65	4.040
	θ	2.29 (1.07)	0.29	1.230	1.94 (0.69)	-0.06	0.478	1.88 (0.62)	-0.12	0.395
	k	95.87 (0.35)	-0.13	0.142	96.31 (0.98)	0.31	1.065	95.72 (0.47)	-0.28	0.295
80	α	1.91 (0.53)	-0.09	0.291	2.55 (1.21)	0.55	1.769	2.49 (1.49)	0.49	2.46
	θ	2.18 (0.85)	0.18	0.747	1.95 (0.65)	-0.05	0.431	1.92 (0.61)	-0.08	0.374
	k	95.90 (0.24)	-0.10	0.066	96.28 (0.94)	0.28	0.968	95.74 (0.48)	-0.26	0.298
100	α	1.90 (0.53)	-0.10	0.286	2.48 (1.16)	0.48	1.578	2.44 (0.25)	0.44	0.255
	θ	2.14 (0.60)	0.14	0.375	1.95 (0.55)	-0.05	0.307	1.92 (0.38)	-0.08	0.152
	k	95.90 (0.21)	-0.10	0.052	96.26 (0.74)	0.26	0.623	95.79 (0.005)	-0.21	0.045
200	α	1.93 (0.28)	-0.07	0.085	2.18 (0.64)	0.18	0.444	2.40 (0.16)	0.40	0.189
	θ	2.08 (0.33)	0.08	0.115	1.98 (0.31)	-0.02	0.099	1.94 (0.19)	-0.06	0.039
	k	95.93 (0.15)	-0.07	0.027	96.10 (0.42)	0.10	0.190	95.85 (0.004)	-0.15	0.021

As the sample size increases for all three censoring scenarios, the bias and mean squared error decreases, verifying the consistency of the estimators (Table 4.1).

4.5 Applications

In this section, we present three real data sets to demonstrate the flexibility and potential of the RSTG distribution in modeling negatively skewed data. We compare the performance of the RSTG distribution to the exponential, generalized F, generalized gamma, Gompertz, log-logistic, lognormal, Rayleigh and Weibull distributions. Density functions of the compared distributions are given in Appendix E. We use four discrimination criteria methods based on the log-likelihood function evaluated at the MLEs. Letting p be the number of parameters to be fitted, n the sample size, and $l(\hat{\alpha}, \hat{\theta}, \hat{k})$ the log-likelihood function, the criteria we consider are the following: Akaike information criteria $AIC = -2l(\hat{\alpha}, \hat{\theta}, \hat{k}) + 2p$, corrected Akaike information criterion $AICC = AIC + \frac{2p(p+1)}{(n-p-1)}$, Hannan-Quinn information criterion $HQIC = -2l(\hat{\alpha}, \hat{\theta}, \hat{k}) + 2p\log(\log(n))$, and the consistent Akaike information criterion $CAIC = -2l(\hat{\alpha}, \hat{\theta}, \hat{k}) + p(\log(n) + 1)$ (Anderson, Burnham, and White, 1998; Hannan and Quinn, 1979). An advantage in using information-theoretic criteria is that it is valid even for non-nested models (Burnham and Anderson, 2002). Burnham and Anderson suggest that an AIC difference of between 3 and 7 units indicates that a candidate model has considerably less support than the model with the minimum AIC value, while a difference of more than 10 units indicates that a candidate model is highly unlikely.

4.5.1 Badenscallie Burial Data

In this first example, ages of death for 59 males members of the Scottish McAlpha clan were collected in June 1987 from the burial ground at Badenscallie in the Coigach district of Wester Ross, Scotland (Sprent and Smeeton, 2007). The ages are recorded

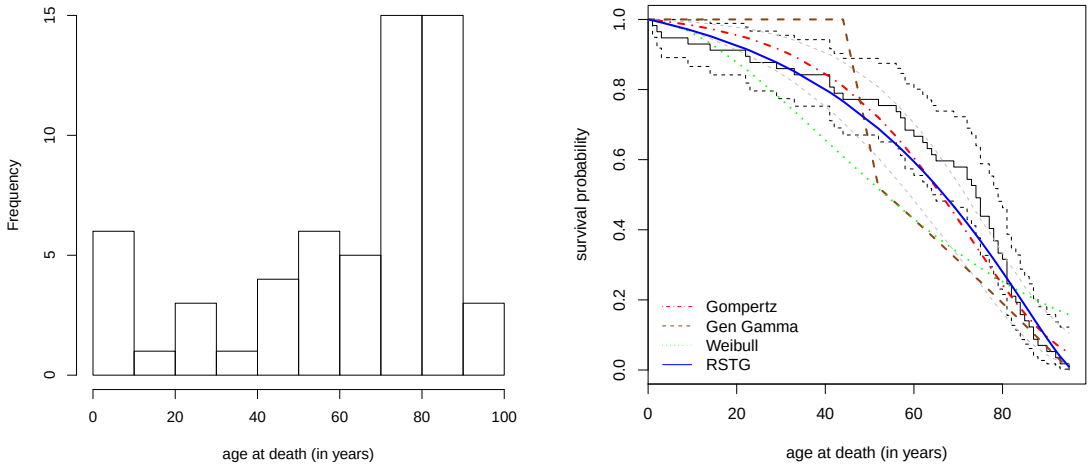

Figure 4.4: Data distribution and survival functions of the four best models of the Badenscallie data.

in complete years, e.g. 0 means before the first birthday and 75 means after the 75^{th} but before the 76^{th} birthday. The negative skew of this distribution is verified by the skewness coefficient of -0.79. The distribution of death times is shown in Figure 4.4.

We compare the performance of the RSTG distribution with the aforementioned distributions. Information theoretic criteria, parameter estimates and standard errors for each of the fitted distributions are given in Table 4.2. A graphical summary of four model fits, relative to the Kaplan-Meier survival curve, is shown in Figure 4.4. Based on the AIC, AICC, HQIC and CAIC values, the RSTG distribution provides a better fit than all other distributions. The RSTG distribution has an AIC value that is more than 10 units lower than the other distributions in Table 4.2 and is thus superior to the compared distributions for modeling negatively skewed complete data.

4.5.2 Diabetic Data

In this second example, survival times were collected for the first 40 diabetic patients enrolled in an Oklahoma Indian diabetes study (Cooray, 2005). This data is a part of a larger sample of 1012 Oklahoma Indians with non-insulin-dependent diabetes

Table 4.2: Model fitting results of the Badenscallie data.

Model	Par	MLE (se)	AIC	AICC	HQIC	CAIC
Exponential	λ	0.016 (0.002)	606.69	606.76	607.50	609.77
Generalized F	μ	4.432 (0.041)	546.99	547.73	550.23	559.30
	σ	0.174 (0.042)				
	Q	4.376 (1.045)				
	P	0.833 (1.515)				
Generalized Gamma	μ	4.550 (0.044)	541.69	542.13	544.12	550.92
	σ	0.133 (0.026)				
	Q	28.3 (161)				
Gompertz	a	0.050 (0.007)	539.68	539.89	541.30	545.84
	b	0.001 (0.0006)				
Log-logistic	α	0.542 (0.121)	625.34	625.55	626.96	631.50
	β	57.74 (0.113)				
Lognormal	μ	3.730 (0.188)	655.37	655.58	656.99	661.53
	σ	1.450 (0.133))				
Rayleigh	b	4.210 (0.065)	592.49	592.56	593.30	595.57
Weibull	λ	1.709 (0.208)	592.49	592.56	593.30	595.57
	γ	66.331 (5.146)				
RSTG	α	1.083 (0.392)	**526.83**	**527.27**	**529.26**	**536.06**
	θ	42.709 (28.261)				
	k	95.065 (0.409)				

mellitus who were examined in 1972 –1980 and had a follow-up study conducted in 1986–1989 (Lee and Wang, 2003). Some of the survival times are right-censored. The skewness coefficient of -1.30 verifies the negative skew of the distribution. The distribution of survival times is shown in Figure 4.5.

We compare the performance of the RSTG distribution with the aforementioned distributions. Information theoretic criteria, parameter estimates and standard errors for each of the fitted distributions are given in Table 4.3. A graphical summary of four model fits, relative to the Kaplan-Meier survival curve, is presented in Figure 4.5. The RSTG distribution has an AIC value that is more than 10 units lower than the other distributions and is thus superior to the compared distributions for modeling negatively skewed right-censored data.

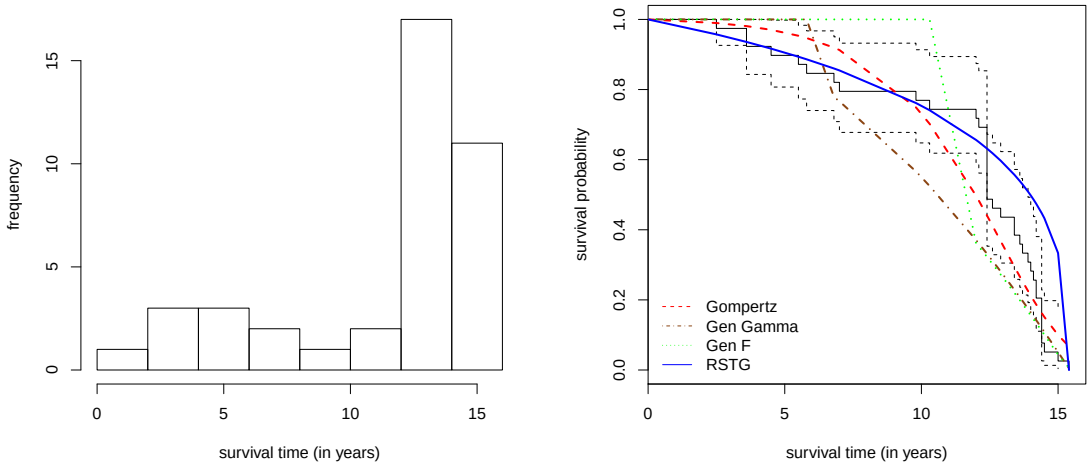

Figure 4.5: Data distribution and survival functions of the four best models of the diabetic data.

4.5.3 Breast Retraction Data

In a final example, times to breast retraction were collected from a retrospective study on 46 early breast cancer patients treated with radiation therapy at the Joint Center for Radiation Therapy in Boston between 1976 and 1980 (Finkelstein and Wolfe, 1985). The breast retraction times are interval-censored. We use the midpoints of each interval-censored observation to calculate a skewness coefficient of -0.71, which verifies the negative skew of the distribution.

The parameter estimates and corresponding standard errors, and the information theoretic criteria for the RSTG, exponential, generalized F, generalized gamma, Gompertz, log-logistic, lognormal, Rayleigh and Weibull distributions are given in Table 4.4. A graphical summary of the RSTG model and the four best models based on the AIC values, relative to the Kaplan Meier survival curve, is shown in Figure 4.6. Based on the information theoretic criteria, the RSTG distribution provides a better fit than the compared distributions.

Table 4.3: Model fitting results of the diabetic data.

Model	Par	MLE(se)	AIC	AICC	HQIC	CAIC
Exponential	λ	0.089 (0.014)	275.56	275.59	276.78	279.56
Generalized F	μ	2.70 (0.005)	215.60	215.88	220.48	231.62
	σ	0.008(0.005)				
	Q	6.87 (4.50)				
	P	0.021(1.32)				
Generalized Gamma	μ	2.73 (0.023)	213.00	213.17	216.66	225.01
	σ	0.025 (0.073)				
	Q	21.7 (62.9)				
Gompertz	a	0.396 (0.060)	210.18	210.26	212.62	218.19
	b	0.002 (0.002)				
Log-logistic	α	2.767 (0.395)	269.39	269.47	271.83	277.40
	β	1.056 (1.014)				
Lognormal	μ	2.191 (0.186)	305.83	305.91	308.27	313.84
	σ	0.886 (0.056)				
Rayleigh	b	2.48 (0.079)	247.93	247.96	249.15	251.93
Weibull	λ	2.561 (0.375)	247.32	247.40	249.76	255.33
	γ	12.177 (0.766)				
RSTG	α	0.232 (0.054)	**149.06**	**149.22**	**152.72**	**161.07**
	θ	46.317 (39.525)				
	k	15.447(0.907)				

4.6 Discussion

The flexibility, applicability and better fit of the RSTG distribution as compared to eight standard distributions has been demonstrated when modeling negatively skewed complete, right-censored and interval-censored survival data by AIC, AICC, HQIC, and CAIC criteria. The data sets used in Examples 4.5.1 and 4.5.2 were previously modeled with extensions of the Gompertz distribution. The Gompertz-sinh family was constructed to model highly negatively skewed survival data with thick lower tails, such as the Badenscallie data used in Example 4.5.1. This data set was analyzed using the Gompertz-sinh and the exponentiated Gompertz-sinh distributions. (Cooray and Ananda, 2010). A logistic-sinh distribution, designed for negatively skewed distributions with long thin tails, has been proposed to model the subset of the diabetic data presented in Example 4.5.2. In addition to its superiority over

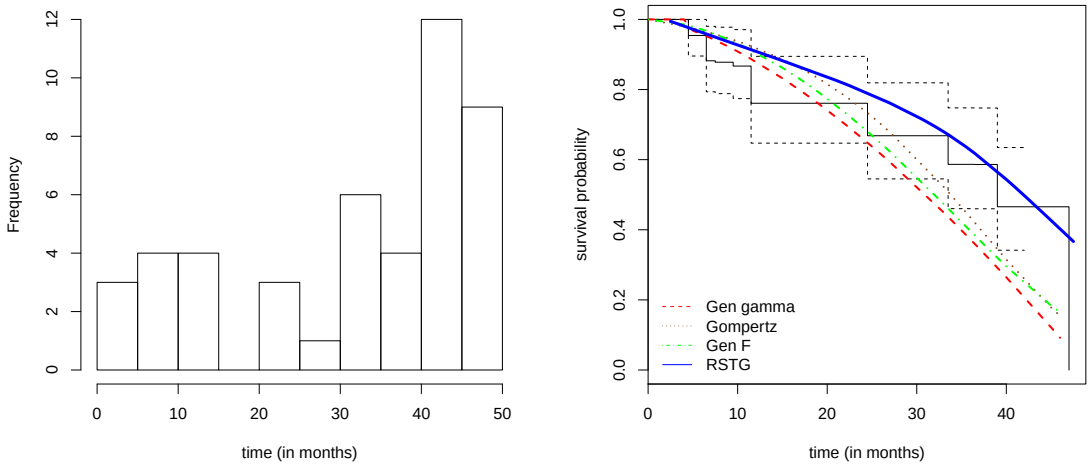

Figure 4.6: Data distribution and survival functions of the four best models of the breast retraction data.

the compared standard distributions, the RSTG distribution is comparable to the Gompertz-sinh family for negatively skewed survival data with thick tails and performs better than the Gompertz-sinh family and the logistic-sinh distribution when modeling negatively skewed distributions with thin tails. The RSTG distribution is a viable alternative when researchers encounter negatively skewed survival data.

Table 4.4: Model fitting results of the breast retraction data.

Model	Par	MLE(se)	AIC	AICC	HQIC	CAIC
Exponential	λ	0.032(0.005)	215.02	215.11	215.73	217.89
Generalized F	μ	3.688(0.080)	185.45	186.38	188.28	196.93
	σ	0.311(0.084)				
	Q	1.628(0.571)				
	P	0.912(1.231)				
Generalized Gamma	μ	3.860(0.022)	162.44	162.99	164.56	171.05
	σ	0.077(0.031)				
	Q	7.777(2.940)				
Gompertz	a	0.077(0.047)	176.58	176.85	177.99	182.32
	b	0.004(0.002)				
Log-logistic	α	2.54(0.115)	209.58	209.85	210.99	215.32
	β	29.21(9.219)				
Lognormal	μ	3.258(0.107)	210.97	211.24	212.38	216.71
	σ	0.706(0.081)				
Rayleigh	b	3.55(0.075)	190.52	190.61	191.23	193.39
Weibull	λ	2.17(0.30)	192.17	192.44	193.58	197.91
	γ	35.33(0.35)				
RSTG	α	1.940(1.325)	**138.98**	**139.25**	**140.39**	**144.72**
	θ	23.234(12.234)				
	k	70.545(16.059)				

Chapter 5

An Accelerated Failure Time Model Using the RSTG Distribution with Application to Pediatric Nephrotic Syndrome

5.1 Introduction

The classical accelerated failure time (AFT) model (Kalbfleisch and Prentice, 2011) provides an attractive alternative to the Cox proportional hazards model (D. R. Cox, 1972) in survival analysis due to its direct physical interpretation. We can express the survival time of one patient as being accelerated or decelerated by some factor as compared to another patient while taking into account covariates that contribute to the change in survival time. Accelerated failure time models allow for a wide range of parametric forms for the survival functions. A fully parametric model has the advantage of a simple framework for maximum likelihood estimation. The parameter estimates then have desirable properties such as asymptotic normality. The suitability of the parametric distribution to the data can easily be assessed using graphical methods, and inference will be far more precise (Collett, 2015).

In Chapter 4, we demonstrated that the RSTG distribution provides a better model fit for negatively skewed complete, right-censored and interval-censored survival data than the exponential, generalized gamma, generalized F, lognormal, log-logistic, Rayleigh, Gompertz and Weibull distributions. In this chapter, we use the RSTG distribution in an accelerated failure time model and apply it to the pediatric nephrotic

syndrome data. The AFT model is our model of choice as we expect that some of the explanatory variables suggested by the literature will actually decelerate the time to remission for the frequently relapsing or steroid dependent nephrotic syndrome patient.

5.2 The Model

The accelerated failure time model has a log-linear representation as

$$\log T_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots\ldots + \beta_p x_{pi} + \sigma\epsilon_i, \tag{5.1}$$

where T_i is the random variable associated with the survival time of the i^{th} individual, β_0 is the true intercept term, $\beta_1, \ldots, \beta_p$ are the regression coefficients of interest, x_{ji} is the j^{th} explanatory variable (covariate) for the i^{th} individual ($i = 1, 2, \ldots, n; j = 1, 2, \ldots, p$), σ is a scalar, and ϵ_i is a random disturbance term assumed to be identically and independently distributed with density function $f(\epsilon_i)$.

The regression coefficients of the model in equation (5.1) are interpreted as follows: If we increase the value of x_{ki} by 1 and hold all other covariates fixed, the change in survival time will increase (if $\beta_k > 0$) or decrease (if $\beta_k < 0$) by a factor of e^{β_k}. In other words, $100e^{\beta_k}$ represents the percentage change in median survival time.

Following the log-linear model formulation in equation (5.1), we express time T as

$$T = e^{x^T\beta + \sigma\epsilon} = e^{x^T\beta}(e^{\epsilon\sigma}) = e^{x^T\beta}(T_0).$$

We assume that $T_0 = e^{\sigma\epsilon}$ follows the RSTG distribution, with density function (4.1), survival function (4.3), and hazard function (4.4).

Given $\Theta = (\alpha, \theta, \boldsymbol{\beta}, k)$, we write the survival function for the RSTG AFT model as

$$\begin{aligned} S_A(t|x,\Theta) &= P(T \geq t) = P(e^{x^T\beta}(T_0) \geq t) \\ &= P(T_0 \geq e^{-x^T\beta}t) \\ &= P(T_0 \geq g(t)) \\ &= \int_{g(t)}^{\infty} \frac{1}{\theta^\alpha \left(\Gamma(\alpha) - \Gamma(\alpha, \frac{k}{\theta})\right)} e^{-\frac{(-x+k)}{\theta}} (-x+k)^{\alpha-1} \mathrm{d}x \\ &= 1 - F^*(g(t)) \\ &= S^*(g(t)) \end{aligned} \tag{5.2}$$

Thus, $S_A(t|x,\Theta) = \frac{\Gamma(\alpha) - \Gamma\left(\alpha, \frac{-g(t)+k}{\theta}\right)}{\Gamma(\alpha) - \Gamma(\alpha, \frac{k}{\theta})}$, where $g(t) = e^{-x^T\beta}t$.

The density function of the RSTG AFT model is

$$\begin{aligned} f_A(t|x,\Theta) &= -\frac{\mathrm{d}(S_A(t))}{\mathrm{d}t} \\ &= -\frac{\mathrm{d}\left(S^*(g(t))\right)}{\mathrm{d}t} \\ &= f^*(g(t)) \cdot g'(t) \\ &= \frac{1}{\theta^\alpha \left(\Gamma(\alpha) - \Gamma(\alpha, \frac{k}{\theta})\right)} e^{-\frac{(-g(t)+k)}{\theta}} (-g(t)+k)^{\alpha-1} \cdot g'(t), \end{aligned}$$

with hazard function

$$\begin{aligned} h_A(t|x,\Theta) = \frac{f_A(t)}{S_A(t)} &= \frac{f^*\left(g(t)\right) \cdot g'(t)}{S^*\left(g(t)\right)} \\ &= h^*\left(g(t)\right) \cdot g'(t) \\ &= \frac{e^{-\frac{(-g(t)+k)}{\theta}} (-g(t)+k)^{\alpha-1}}{\theta^\alpha \left(\Gamma(\alpha) - \Gamma\left(\alpha, \frac{k}{\theta}\right)\right)} \cdot g'(t), \end{aligned}$$

where $g'(t) = e^{-x^T\beta}$.

Without loss of generality, we use $S_A(t)$, $f_A(t)$ and $h_A(t)$ to represent $S_A(t|x,\Theta)$, $f_A(t|x,\Theta)$ and $h_A(t|x,\Theta)$.

5.3 Parametric Estimation

Several approaches have been proposed for the estimation and inference of the semi-parametric AFT model. One method involves rank-based estimators as first discussed in the literature by Prentice (1978), and another is the method of Buckley and James (1979), which provides an accommodation of the least-squares estimator. The asymptotic properties of the two estimators have been studied by many authors (Tsiatis, 1990; Ritov, 1990; Jin, Lin, Wei, and Ying, 2003). Classically, the estimation of the parameters in a fully parametric AFT model is performed by maximizing the log-likelihood equation (David Roxbee Cox and Oakes, 1984; Robins, 1992). We will use maximum likelihood methods for estimation of the parameters of the RSTG AFT model.

For data that contains both complete and right-censored information, the likelihood function of the RSTG AFT model for the parameter vector $\Theta = (\alpha, \theta, \boldsymbol{\beta}, k)$ is proportional to

$$\begin{aligned} L(\Theta) &= \prod_{i=1}^{n} f_A(t_i)^{\delta_i} S_A(t_i)^{1-\delta_i} \\ &= \prod_{i=1}^{n} \left[\frac{1}{\theta^\alpha \left(\Gamma(\alpha) - \Gamma(\alpha, \frac{k}{\theta}) \right)} e^{-\frac{(-g(t_i)+k)}{\theta}} (-g(t_i)+k)^{\alpha-1} \cdot g'(t_i) \right]^{\delta_i} \\ &\times \left[\frac{\Gamma(\alpha) - \Gamma\left(\alpha, \frac{-g(t_i)+k}{\theta}\right)}{\Gamma(\alpha) - \Gamma\left(\alpha, \frac{k}{\theta}\right)} \right]^{1-\delta_i}, \end{aligned}$$

where $g(t) = e^{-x^T\beta}t$. The censoring indicator δ_i is such that

$$\delta_i = \begin{cases} 1 & \text{if the event of interest is observed} \\ 0 & \text{if the event of interest is not observed (event time is right-censored)} \end{cases}.$$

The corresponding log-likelihood function is given by

$$
\begin{aligned}
l(\Theta) &= \ln L(\Theta) \\
&= \sum_{i=1}^{n} \delta_i \Bigg\{ - \left[\alpha \ln \theta + \ln \left(\Gamma(\alpha) - \Gamma\left(\alpha, \frac{k}{\theta}\right)\right)\right] - \frac{-g(t_i)+k}{\theta} \\
&\quad + (\alpha - 1) \ln\left(-g(t_i)+k\right) + \ln g'(t_i) \Bigg\} \\
&\quad + \sum_{i=1}^{n} (1-\delta_i) \left\{ \ln \left[\Gamma(\alpha) - \Gamma\left(\alpha, \frac{-g(t_i)+k}{\theta}\right)\right] - \ln \left[\Gamma(\alpha) - \Gamma\left(\alpha, \frac{k}{\theta}\right)\right] \right\}.
\end{aligned}
\tag{5.3}
$$

We take partial derivatives of equation (5.3) with respect to α, θ, k, and each β_j to obtain the normal equations for the unknown parameters and equate each to zero. Without loss of generality, we let $l = l(\Theta)$. The resulting equations are

$$
\begin{aligned}
\frac{dl}{d\alpha} &= \sum_{i=1}^{n} \delta_i \left(- \ln\theta - \frac{\Gamma'(\alpha) - \Gamma'\left(\alpha, \frac{k}{\theta}\right)}{\Gamma(\alpha) - \Gamma\left(\alpha, \frac{k}{\theta}\right)} + \ln(-g(t_i)+k) \right) \\
&\quad + \sum_{i=1}^{n} (1-\delta_i) \left(\frac{\Gamma'(\alpha) - \Gamma'\left(\alpha, \frac{-g(t_i)+k}{\theta}\right)}{\Gamma(\alpha) - \Gamma\left(\alpha, \frac{-g(t_i)+k}{\theta}\right)} - \frac{\Gamma'(\alpha) - \Gamma'\left(\alpha, \frac{k}{\theta}\right)}{\Gamma(\alpha) - \Gamma\left(\alpha, \frac{k}{\theta}\right)} \right) = 0,
\end{aligned}
\tag{5.4}
$$

where $\Gamma'(\alpha) = \frac{d\Gamma(\alpha)}{d\alpha} = \frac{d}{d\alpha}(\ln\Gamma(\alpha))\Gamma(\alpha)$, and $\Gamma'(\alpha, \frac{k}{\theta}) = \frac{d\Gamma(\alpha, \frac{k}{\theta})}{d\alpha} = \int_{\frac{k}{\theta}}^{\infty} \ln(y) y^{\alpha-1} e^{-y} \mathrm{d}y$,

$$
\begin{aligned}
\frac{dl}{d\theta} &= \sum_{i=1}^{n} \delta_i \left(-\frac{\alpha}{\theta} + \frac{\frac{1}{\theta}\left(\frac{k}{\theta}\right)^{\alpha} e^{-\frac{k}{\theta}}}{\Gamma(\alpha) - \Gamma\left(\alpha, \frac{k}{\theta}\right)} + \frac{-g(t_i)+k}{\theta^2} \right) \\
&\quad + \sum_{i=1}^{n} (1-\delta_i) \left(\frac{-\frac{1}{\theta}\left(\frac{-g(t_i)+k}{\theta}\right)^{\alpha} e^{-\frac{-g(t_i)+k}{\theta}}}{\Gamma(\alpha) - \Gamma\left(\alpha, \frac{-g(t_i)+k}{\theta}\right)} + \frac{\frac{1}{\theta}\left(\frac{k}{\theta}\right)^{\alpha} e^{-\frac{k}{\theta}}}{\Gamma(\alpha) - \Gamma\left(\alpha, \frac{k}{\theta}\right)} \right) = 0,
\end{aligned}
\tag{5.5}
$$

$$
\begin{aligned}
\frac{dl}{dk} &= \sum_{i=1}^{n} \delta_i \left(\frac{-\frac{1}{k}\left(\frac{k}{\theta}\right)^{\alpha} e^{-\frac{k}{\theta}}}{\Gamma(\alpha) - \Gamma\left(\alpha, \frac{k}{\theta}\right)} - \frac{1}{\theta} + \frac{(\alpha-1)}{-g(t_i)+k} \right) \\
&\quad + \sum_{i=1}^{n} (1-\delta_i) \left(\frac{\frac{1}{-g(t_i)+k}\left(\frac{-g(t_i)+k}{\theta}\right)^{\alpha} e^{-\frac{-g(t_i)+k}{\theta}}}{\Gamma(\alpha) - \Gamma\left(\alpha, \frac{-g(t_i)+k}{\theta}\right)} - \frac{\frac{1}{k}\left(\frac{k}{\theta}\right)^{\alpha} e^{-\frac{k}{\theta}}}{\Gamma(\alpha) - \Gamma\left(\alpha, \frac{k}{\theta}\right)} \right) = 0,
\end{aligned}
\tag{5.6}
$$

and

$$\begin{aligned}\frac{dl}{d\beta_j} &= \sum_{i=1}^{n}\delta_i\left(\frac{-x_j e^{-x^T\beta}t_i}{\theta}+(\alpha-1)\frac{x_j e^{-x^T\beta}t_i}{-e^{-x^T\beta}t_i+k}-x_j\right)\\ &+\sum_{i=1}^{n}(1-\delta_i)\left\{-x_j e^{-x^T\beta}t_i\left(\frac{1}{\theta}\left(-e^{-x^T\beta}t_i+k\right)\right)^{\alpha}\right.\\ &\left.\times\frac{e^{\frac{1}{\theta}\left(e^{-x^T\beta}t_i-k\right)}}{\Gamma(\alpha)\left(e^{-x^T\beta}t_i-k\right)+\Gamma\left(\alpha,\frac{1}{\theta}\left(-e^{-x^T\beta}t_i+k\right)\right)\left(-e^{-x^T\beta}t_i+k\right)}\right\}=0.\end{aligned} \tag{5.7}$$

Techniques discussed in Section 4.3.1 can be applied to equations (5.4), (5.5), (5.6), and (5.7) to find the MLEs. These equations cannot be solved explicitly. We use iterative methods in **R** to obtain the MLEs and standard errors.

5.4 Application to Pediatric Nephrotic Syndrome

An illustration of the superior performance of the proposed RSTG AFT regression model (5.1) compared to AFT regression models with other distributional assumptions is given in Appendix F. The comparisons are based on information theoretic criteria and standard errors of the parameter estimates. We now apply the RSTG AFT model to the pediatric nephrotic syndrome data from HSSC and from the Robert Wood Johnson Medical School.

5.4.1 HSSC Data

To demonstrate the applicability of the RSTG distribution to real negatively skewed data, we first analyze pediatric nephrotic syndrome data from the HSSC CDW described in Chapter 3.

Hospital readmission, particularly in the pediatric population, has been the focus of previous studies (Feudtner et al., 2009; Gay, Hain, Grantham, and Saville, 2011). The times to the first hospital visit within the first 30 days after diagnosis of pediatric NS were recorded from a retrospective analysis of the HSSC data. We choose 30 days as a censoring point to account for possible scheduled return visits, and because previous

literature suggests that only a small percentage of pediatric nephrotic syndrome patients have not entered remission after 4 weeks (Constantinescu et al., 2000). Previous literature also suggests that the time to remission for the frequently relapsing or steroid dependent nephrotic syndrome patient is longer than that of the infrequently relapsing patient (Constantinescu et al., 2000; Nakanishi et al., 2013). According to one American study, 52% of IFRNS patients achieved initial remission by 7 days after diagnosis, while only 21% of the FRNS and SDNS patients achieved initial remission by 7 days ($\chi^2 = 4.5; p = 0.03$) (Constantinescu et al., 2000). We hypothesize that those whose time to initial remission is longer, i.e. the FRNS or SDNS patient, are more likely to have a return hospital visit within a 30-day period. This return visit could be prompted by complications arising from the body being in the state of the nephrotic syndrome, such as severe edema, hypertension, or bacterial peritonitis (Richardson, 2012; Debbie S. Gipson et al., 2009). The return visit could also occur because of adverse events resulting from the prolonged use of high dosage corticosteroid therapy which is classically used for the initial diagnosis of idiopathic pediatric nephrotic syndrome.

We use the continuous covariate age and categorical 0/1 covariates representing the season of diagnosis (with spring being the referent level) to identify predictive factors for the time to the first hospital visit after diagnosis of the nephrotic syndrome. We define season of diagnosis as: fall (August, September, October); winter (November, December, January); spring (February, March, April) and summer (May, June, July). The distribution of times to first hospital visit after diagnosis has skewness coefficient -0.88 (Figure 5.1).

The AFT model for this data is given by

$$\log(T_i) = \beta_0 + \beta_1 age_i + \beta_2 winter_i + \beta_3 summer_i + \beta_4 fall_i + \sigma\epsilon_i. \tag{5.8}$$

where T_i represents the time to first hospital visit for the i^{th} patient, $i = 1, .., n$.

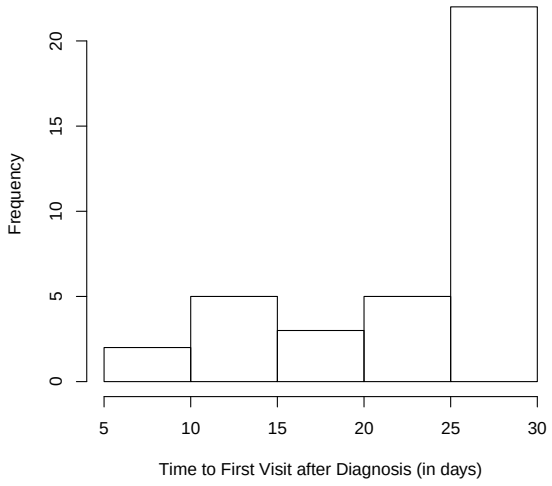

Figure 5.1: Distribution of times to first hospital visit after initial diagnosis of pediatric nephrotic syndrome from HSSC CDW.

We analyze the data under the accelerated failure time framework with different distributional assumptions. Results are presented in Appendix F.

To assess the suitability of a model, graphical checks may be preferred over formal statistical tests of lack of fit because the formal tests tend to have low power for small-sample sizes or they always reject a given model for large sample sizes (Klein and Moeschberger, 2003). Cox-Snell residuals, Martingale residuals and deviance residuals are three types of residuals that are commonly used to assess the fit of a model graphically (see John P. Klein and Moeschberger (2003) and Collett (2015) for a discussion of each). We evaluate the accuracy of the RSTG AFT model using a diagnostic plot of the deviance residuals. The deviance residuals, first introduced by Terry M. Therneau, Grambsch, and Fleming (1990), can be expected to be symmetrically distributed about zero when an appropriate model has been fit. The deviance residuals are defined as $r_{D_i} = sgn(M_i)[-2\{M_i + \delta_i \log(\delta_i - M_i)\}]^{\frac{1}{2}}$, where δ_i is the censoring indicator, $M_i = \delta_i + \log \hat{S}_A(t_i)$, and $sgn(\cdot)$ is a function that simply

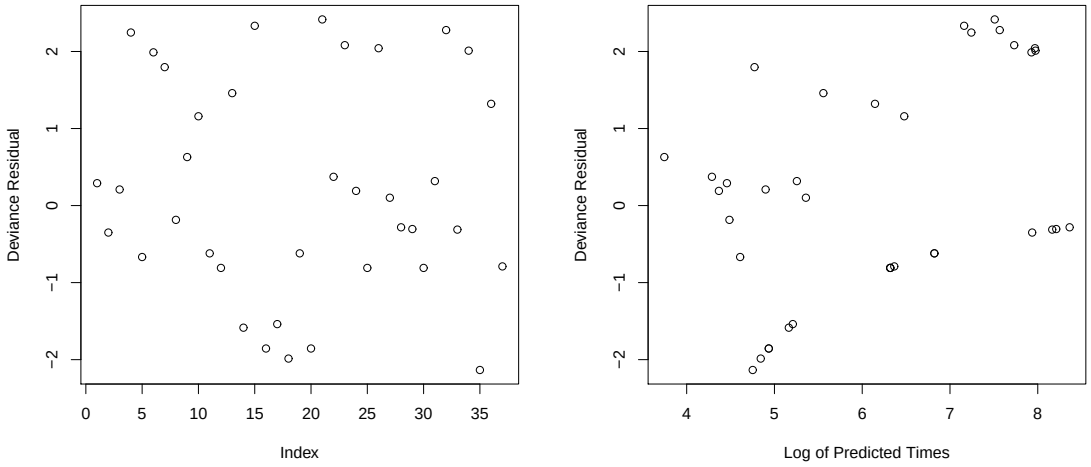

Figure 5.2: Deviance residual plots of the RSTG AFT model for the HSSC CDW pediatric nephrotic syndrome data.

Table 5.1: RSTG AFT model for pediatric nephrotic syndrome patients in South Carolina.

Distribution	Parameter	MLE(se)
RSTG	β_1	0.046 (0.060)
	β_2	3.153 (2.820)
	β_3	1.384 (1.499)
	β_4	3.138 (4.420)

takes the sign of the argument. A plot of the deviance residuals indicates that the RSTG AFT model provides a good fit for the data (Figure 5.2).

No significant predictors of the time to first hospital visit after diagnosis were detected by the RSTG AFT model (Table 5.1).

5.4.2 Robert Wood Johnson Medical School Data

As a second example demonstrating the applicability of the RSTG distribution to negatively skewed data, we use the New Jersey data described in Chapter 3. Previous analyses of the New Jersey data reported a median time to initial remission of 7 days for IFRNS patients and a median time to initial remission of more than 7 days for

the FRNS and SDNS patients (Constantinescu et al., 2000). These median remission times coincide with those of a later study by Vivarelli et al. (2010). The original study of this data analyzed odds ratios between the IFRNS group and the FRNS/SDNS group. Each group was assessed for time to first remission with censoring after 7 days (Constantinescu et al., 2000). A significant association was found between initial remission times less than 7 days and an IFR diagnosis for those patients who did not have hematuria at diagnosis. Study authors also report that they did not take into account the histopathology found on renal biopsy.

The objective of our study of the New Jersey data is to identify early prognostic factors for idiopathic nephrotic syndrome, particularly FRNS or SDNS. We use the RSTG AFT model to examine predictors of the time to first remission for pediatric nephrotic syndrome patients. The variables age at diagnosis, hematuria (0 = not present at diagnosis, 1 = present at diagnosis), and creatinine level (mg/dL) are used to determine their effects on the accelerated or decelerated time to initial remission. All patients were initially treated with the standard corticosteroid therapy. Following the ideas of the original study authors, we fit the model using censoring at 7 days after diagnosis. Censoring at 7 days results in a negatively skewed distribution of initial remission times with skewness coefficient -2.68.

The AFT model for this data is given by

$$\log(T_i) = \beta_0 + \beta_1 age_i + \beta_2 hematuria_i + \beta_3 creatinine_i + \sigma\epsilon_i, \tag{5.9}$$

where T_i represents the time to first remission for the i^{th} patient, $i = 1, .., n$. Censoring at 7 days for the RSTG AFT model did not detect significance of any predictors of the time to first remission ($\beta_1 = -0.012, SE = 0.026; \beta_2 = -0.330, SE = 0.201; \beta_3 = 0.102, SE = 0.562$). We explore later censoring times to determine if any significant effects are present.

Further review of the New Jersey data reveals that the mean time to first remission for FRNS and SDNS patients is 10 days and the median time is 11.5 days. Study

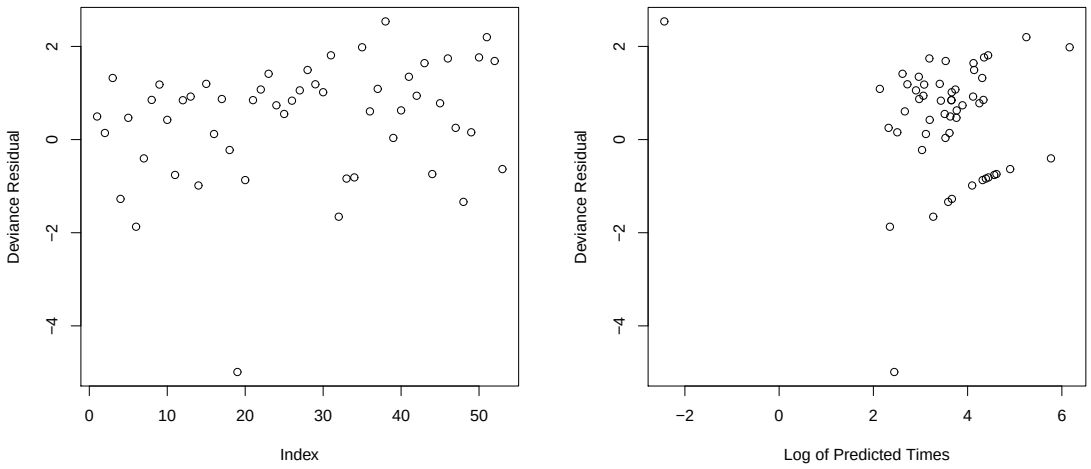

Figure 5.3: Deviance residual plots of the RSTG AFT model for the New Jersey pediatric nephrotic syndrome data.

findings also indicate that 75% of IFRNS patients are in remission at 14 days. Because the objective is to identify as early as possible significant predictors of NS patients who tend toward a FR or SD course, we use the study the data for up to 14 days after diagnosis. This is consistent with a startup study with follow-up time of 14 days. This convention results in a negatively skewed distribution of initial remission times with skewness coefficient -0.40. We apply the RSTG AFT model to the data. A plot of the deviance residuals indicates the presence of outliers that may affect model fit (Figure 5.3). One outlier corresponds to an individual who entered spontaneous remission while the other corresponds to an individual with a very low creatinine level at diagnosis. The model was refit without the outliers and neither the magnitude nor significance of the parameter estimates changed substantially. Original model fitting results are given in Table 5.2.

Based on the parameter estimates and standard errors, the model suggests that age at diagnosis is a significant indicator of the time to first remission ($\beta_1 = -0.172, SE = 0.052$). After controlling for hematuria status and creatinine level, the time to first remission for pediatric NS patients decreases by 16%(95% CI : $7\% - 24\%$) for each one

Table 5.2: RSTG AFT model for pediatric nephrotic syndrome patients in New Jersey.

Distribution	Parameter	MLE(se)
RSTG	β_1	-0.172 (0.052)
	β_2	0.396 (0.592)
	β_3	5.158 (3.245)

year increase in age at diagnosis. This finding supports the findings of R. F. Andersen et al. (2010), who suggested that early age at debut in a significant predictor of SDNS and FRNS.

5.5 Discussion

Potential two-way interactive effects were explored in the Robert Wood Johnson medical school data. No interactions were found to be significant.

Chapter 6

Frailty Models using the RSTG Distribution with Application to Pediatric Nephrotic Syndrome and Diabetic Retinopathy

6.1 Introduction

In practice, there can exist either unobserved or unmeasurable effects that cause an individual or group of individuals to experience an event sooner or later than expected. These non-measured random effects, commonly referred to as frailties, may evoke significant changes in survival probabilities if accounted for in the modeling process. The concept of frailty was introduced as early as 1979 in a discussion of the impact of heterogeneity of individual frailty on the dynamics of mortality (Vaupel, Manton, and Stallard, 1979). The underlying logic of frailty models is that some subjects (or groups or clusters) are intrinsically more or less prone to experience an event of interest than are others, and that the distribution of these effects can be at least approximated (Box-Steffensmeier and De Boef, 2006).

Frailty models in survival analysis are commonly used to quantify the association between individual survival times within a subgroup (John P. Klein and Moeschberger, 2003). For example, there may be a shared genetic structure or a shared environmental factor that leads to dependence among the event times. This shared frailty concept was first introduced by Clayton (1978) and has been studied more extensively by many researchers, most within the framework of the PH model (Oakes, 1989; McGilchrist

and Aisbett, 1991; Hougaard, 1986b; John P Klein, Moeschberger, Li, Wang, and Flournoy, 1992; Keiding, Andersen, and Klein, 1997; John P. Klein and Moeschberger, 2003; Terry M Therneau, Grambsch, and Pankratz, 2003; Chen et al., 2013). While the shared frailty model accounts for unobserved covariates that operate at categorized levels above the individual unit, the individual frailty model accounts for heterogeneity among individual units (Gutierrez et al., 2002). The individual frailty model can be used to model the effect of important covariates that have not been observed (Wienke, 2010). The correlated frailty model, in which the frailties of individuals in a cluster are correlated but not necessarily shared, is a natural extension of both the individual and the shared frailty model concept (Wienke, 2010). Correlated frailty models have been used in multiple studies, including studies of diabetic retinopathy (Huster, Brookmeyer, and Self, 1989), studies of acquisition of both Hepatitis A and Hepatitis B (Hens, Wienke, Aerts, and Molenberghs, 2009), and studies of kidney infection (Hanagal, Pandey, and Ganguly, 2015). Frailty models can also be used to model event dependence arising from repeated occurrence of the same type of event within an individual. Examples include recurrent hospitalizations for transplant candidates with kidney disease, pulmonary exacerbations in cystic fibrosis asthma attacks, or relapse of diseases (Greenwood and Yule, 1920; Box-Steffensmeier and De Boef, 2006; L. Liu, Wolfe, and Huang, 2004; Oakes, 1992).

The choice of frailty distribution plays an important role in the survival model. Theoretically, any non-negative distribution can be used as a frailty distribution. The most commonly used distributions are the gamma and the lognormal, but others include the inverse Gaussian, inverse gamma and the positive stable distribution (Aalen, 1994; P. K. Andersen, Klein, Knudsen, and y Palacios, 1997; Balakrishnan and Peng, 2006; Duchateau and Janssen, 2007; Wienke, 2010; Hougaard, 1986, 2012). The choice of the frailty distribution is often driven by mathematical convenience (Chen et al., 2013). The effects of different frailty distributions have been investigated by several

authors, including Pickles and Crouchley (1995) and Hanagal and Sharma (2015). The use of gamma distributed frailty in univariate survival models is supported by the results of Abbring and Van Den Berg (2007), who showed that, under some regularity assumptions, frailty among survivors converges against a gamma distribution even if the original distribution is not a gamma distribution.

The use of frailties in the AFT framework has seen increased usage by researchers over the past fifteen years. These researchers include Pan (2001), Lambert et al. (2004), Zhang and Peng (2007), and Chen et al. (2013). In this chapter, we investigate the performance of the individual frailty AFT model using the RSTG distribution as the baseline survival distribution with a gamma frailty distribution. An expectation-maximization (EM) algorithm is used for parameter estimation. We apply the algorithm to an individual frailty model using the New Jersey pediatric nephrotic syndrome data. A brief example of the applicability of the RSTG distribution in the correlated frailty model is also presented. We use the correlated frailty model on the 1972 Diabetic Retinopathy study data and compare findings to previously published findings from this data.

6.2 The Frailty Model

The log-linear formulation of the accelerated failure time model is written as

$$\log T_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \ldots\ldots + \beta_p x_{pij} + \sigma\epsilon_{ij},$$

where T_{ij} is the random variable associated with the survival time of the j^{th} individual in the i^{th} cluster. We introduce ω_i to represent either an individual random effect or the random effect shared by individuals in the i^{th}cluster. The new model is expressed as

$$\log T_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \ldots\ldots + \beta_p x_{pij} + \omega_i + \sigma\epsilon_{ij}. \tag{6.1}$$

Following the log-linear model formulation in equation (6.1), we express time T as

$$\begin{aligned} T &= e^{x^T\beta+\omega+\sigma\epsilon} \\ &= e^{x^T\beta} * e^{\omega} * (e^{\epsilon\sigma}) \\ &= e^{x^T\beta}e^{\omega}(T_0) \end{aligned} \tag{6.2}$$

where $z = e^{\omega}$ is the multiplicative frailty term. A frailty value greater than one implies an increased hazard of the event of interest occurring while a frailty value less than one implies a decreased hazard of the event of interest occurring. The variance of the frailty distribution summarizes the degree of heterogeneity among clusters (John P. Klein and Moeschberger, 2003).

We assume that $T_0 = e^{\sigma\epsilon}$ follows the reflected-shifted-truncated-gamma distribution, with density $f^*(t|\alpha,\theta,k)$ given by equation (4.1), survival function $S^*(t|\alpha,\theta,k)$ given by equation (4.3) and hazard function $h^*(t|\alpha,\theta,k)$ given by equation (4.4).

Given the parameter vector $\Theta = (\alpha,\theta,\boldsymbol{\beta},k)$, we express the survival function for the AFT RSTG frailty model as

$$\begin{aligned} S^A(t|x,\Theta) &= P(T \geq t) = P(e^{x^T\beta}e^{\omega}(T_0) \geq t) \\ &= P(T_0 \geq e^{-x^T\beta}e^{-\omega}t) \\ &= P(T_0 \geq g(t)) \\ &= \int_{g(t)}^{\infty} \frac{1}{\theta^{\alpha}\left(\Gamma(\alpha) - \Gamma(\alpha,\frac{k}{\theta})\right)} e^{-\frac{(-x+k)}{\theta}}(-x+k)^{\alpha-1}\mathrm{d}x \\ &= 1 - F^*(g(t)) \\ &= S^*(g(t)) \end{aligned}$$

Thus, $S^A(t|x,\Theta) = \frac{\Gamma(\alpha)-\Gamma\left(\alpha,\frac{-g(t)+k}{\theta}\right)}{\Gamma(\alpha)-\Gamma(\alpha,\frac{k}{\theta})}$, where $g(t) = e^{-x^T\beta}e^{-\omega}t$.

The density function of the AFT RSTG frailty model is

$$\begin{aligned} f^A(t|x,\Theta) &= -\frac{\mathrm{d}(S^A(t))}{\mathrm{d}t} \\ &= -\frac{\mathrm{d}\left(S^*(g(t))\right)}{\mathrm{d}t} \\ &= f^*(g(t)) \cdot g'(t) \\ &= \frac{1}{\theta^\alpha \left(\Gamma(\alpha) - \Gamma(\alpha, \frac{k}{\theta})\right)} e^{-\frac{(-g(t)+k)}{\theta}} (-g(t)+k)^{\alpha-1} \cdot g'(t), \end{aligned}$$

with hazard function

$$\begin{aligned} h^A(t|x,\Theta) = \frac{f^A(t)}{S^A(t)} &= \frac{f^*\left(g(t)\right) \cdot g'(t)}{S^*\left(g(t)\right)} \\ &= h^*\left(g(t)\right) \cdot g'(t) \\ &= \frac{e^{-\frac{(-g(t)+k)}{\theta}} (-g(t)+k)^{\alpha-1}}{\theta^\alpha \left(\Gamma(\alpha) - \Gamma\left(\alpha, \frac{k}{\theta}\right)\right)} \cdot g'(t), \end{aligned}$$

where $g'(t) = e^{-x^T\beta} e^{-\omega}$.

6.3 Parametric Estimation

For a fully parametric model with right-censored observations and no random effects, the likelihood function of the parameter vector Θ for the observations in the i^{th} cluster is proportional to

$$L_i(\Theta) = \prod_{j=1}^{n_i} f^A(t_{ij})^{\delta_i} S^A(t_{ij})^{1-\delta_i}$$

The censoring indicator δ_i is such that

$$\delta_i = \begin{cases} 1 & \text{if the event of interest is observed} \\ 0 & \text{if the event of interest is not observed (event time is right-censored)} \end{cases}.$$

For the model with random effects ω_i, the effects are not known but are assumed to be independent and identically distributed realizations of a random variable with probability density $f(\omega_i)$. The complete likelihood function can be written as

$$L(\Theta) = \prod_{i=1}^{n} \prod_{j=1}^{n_i} f^A(t_{ij})^{\delta_i} S^A(t_{ij})^{1-\delta_i} f(\omega_i)$$

Where n is the number of clusters and n_i is the number of elements in the i^{th} cluster.

We assume gamma frailty for the model, specifically $z_i = e^{\omega_i} \sim \Gamma(\lambda, \frac{1}{\lambda})$. The density function of z is given by $f(z) = \frac{1}{(\frac{1}{\lambda})^\lambda \Gamma(\lambda)} z^{\lambda-1} e^{-\lambda z}$. This distribution has mean 1 and variance $\frac{1}{\lambda}$.

The density function of $\omega_i = \ln z_i$ is

$$\begin{aligned} f(\omega_i) &= \frac{e^{\omega_i(\lambda-1)} e^{-\lambda e^{\omega_i}}}{(\frac{1}{\lambda})^\lambda \Gamma(\lambda)} e^{\omega_i} \\ &= \frac{e^{\lambda\omega_i} e^{-\lambda e^{\omega_i}}}{(\frac{1}{\lambda})^\lambda \Gamma(\lambda)}. \end{aligned}$$

The likelihood function of the RSTG AFT frailty model is then proportional to

$$\begin{aligned} L(\Theta) = \prod_{i=1}^{n} \prod_{j=1}^{n_i} &\left[\frac{1}{\theta^\alpha \left(\Gamma(\alpha) - \Gamma(\alpha, \frac{k}{\theta}) \right)} e^{-\frac{(-g(t_{ij})+k)}{\theta}} (-g(t_{ij}) + k)^{\alpha-1} \cdot g'(t_{ij}) \right]^{\delta_{ij}} \cdot \\ &\left[\frac{\Gamma(\alpha) - \Gamma\left(\alpha, \frac{-g(t_{ij})+k}{\theta}\right)}{\Gamma(\alpha) - \Gamma\left(\alpha, \frac{k}{\theta}\right)} \right]^{1-\delta_{ij}} \cdot \frac{e^{\lambda\omega_i} e^{-\lambda e^{\omega_i}}}{(\frac{1}{\lambda})^\lambda \Gamma(\lambda)} \end{aligned} \tag{6.3}$$

where $g(t_{ij}) = e^{-x_{ij}^T \beta} e^{-\omega_i} t_{ij}$ and $g'(t_{ij}) = e^{-x_{ij}^T \beta} e^{-\omega_i}$.

The corresponding log-likelihood function is given by

$$\begin{aligned} l(\Theta) = \ln L(\Theta) = \sum_{i=1}^{n} \Bigg\{ \sum_{j=1}^{n_i} \delta_{ij} \Bigg\{ &- \left[\alpha \ln\theta + \ln\left(\Gamma(\alpha) - \Gamma\left(\alpha, \frac{k}{\theta}\right) \right) \right] \\ &- \frac{-g(t_{ij})+k}{\theta} + (\alpha-1)\ln(-g(t_{ij})+k) + \ln(g'(t_{ij})) \Bigg\} \\ &+ (1-\delta_{ij}) \Bigg\{ \ln\left[\Gamma(\alpha) - \Gamma\left(\alpha, \frac{-g(t_{ij})+k}{\theta}\right) \right] - \ln\left[\Gamma(\alpha) - \Gamma\left(\alpha, \frac{k}{\theta}\right) \right] \\ &+ \lambda\omega_i - \lambda e^{\omega_i} - \lambda \ln\left(\frac{1}{\lambda}\right) - \ln\Gamma(\lambda) \Bigg\} \Bigg\}. \end{aligned} \tag{6.4}$$

For ease of computation, we write the function $l(\Theta) = l_1(\alpha, \theta, k, \lambda) + l_2(\beta, \alpha, \theta, k, \lambda)$ where

$$l_1(\alpha, \theta, k, \lambda) = \sum_{i=1}^{n} \left\{ \sum_{j=1}^{n_i} \delta_{ij} \left\{ - \left[\alpha \ln \theta + \ln \left(\Gamma(\alpha) - \Gamma\left(\alpha, \frac{k}{\theta}\right)\right)\right] \right\} \right.$$
$$\left. - (1 - \delta_{ij}) \left\{ \ln \left[\Gamma(\alpha) - \Gamma\left(\alpha, \frac{k}{\theta}\right)\right] + \lambda\omega_i - \lambda e^{\omega_i} - \lambda \ln\left(\frac{1}{\lambda}\right) - \ln \Gamma(\lambda) \right\} \right\}.$$

and

$$l_2(\boldsymbol{\beta}, \alpha, \theta, k, \lambda) = \sum_{i=1}^{n} \left\{ \sum_{j=1}^{n_i} \delta_{ij} \left\{ - \frac{-g(t_{ij}) + k}{\theta} + (\alpha - 1) \ln\left(-g(t_{ij}) + k\right) + \ln\left(g'(t_{ij})\right) \right\} \right.$$
$$\left. + (1 - \delta_{ij}) \ln \left[\Gamma(\alpha) - \Gamma\left(\alpha, \frac{-g(t_{ij}) + k}{\theta}\right)\right] \right\}.$$

Without loss of generality, we write $l_1(\alpha, \theta, k, \lambda)$ as l_1 and $l_2(\boldsymbol{\beta}, \alpha, \theta, k, \lambda)$ as l_2.

Due to the unknown random variable ω_i, we cannot maximize the logarithm of the complete likelihood function directly. Following methods similar to those employed by Chen et al. (2013), we use an EM algorithm. The EM algorithm, first introduced by Dempster, Laird, and Rubin in 1977, is an iterative optimization algorithm that alternates between an expectation step (E-step) and a maximization step (M-step). It is a popular approach for finding maximum likelihood estimates of parameters in statistical models that depend on unobserved or unknown random variables.

E-STEP:

The E-step will calculate the conditional expectation of the log-likelihood with respect to the conditional distribution of the random variable ω_i, given the observed data and the estimates of the parameters. We use equation (6.3) and Bayes' Theorem to find the posterior density of ω_i. The probability of ω_i conditional on t_i is

$$\pi(\omega_i | t_i) = \frac{L(t_i|\omega_i) f(\omega_i)}{P(t_i)},$$

where $L(t_i|\omega_i)$ represents the likelihood of the i^{th} event for a fixed ω_i, $f(\omega_i)$ is the probability of a given value of ω_i and $P(t_i)$ is the marginal probability of the data

obtained by integrating $L(t_i|\omega_i)f(\omega_i)$ with respect to ω_i (Collett, 2015) . Ignoring terms in $L(t_i|\omega_i)$ that do not involve ω_i, the posterior density of ω_i is proportional to

$$\prod_{j=1}^{n_i}\left(e^{-\frac{(-g(t_{ij})+k)}{\theta}}(-g(t_{ij})+k)^{\alpha-1}\cdot g'(t_{ij})\right)^{\delta_{ij}}\left(\Gamma(\alpha)-\Gamma\left(\alpha,\frac{-g(t_{ij})+k}{\theta}\right)\right)^{1-\delta_{ij}}$$
$$\times\, e^{\lambda\omega_i}e^{-\lambda e^{\omega_i}}. \tag{6.5}$$

The conditional expectation of $l(\Theta)$ can be written as $\mathrm{E}(l(\Theta)) = \mathrm{E}(l_1) + \mathrm{E}(l_2)$, where

$$\mathrm{E}(l_1) = \sum_{i=1}^{n}\left\{\sum_{j=1}^{n_i}\delta_{ij}\left\{-\left[\alpha\ln\theta+\ln\left(\Gamma(\alpha)-\Gamma\left(\alpha,\frac{k}{\theta}\right)\right)\right]\right\}\right.$$
$$\left.-(1-\delta_{ij})\left\{\ln\left[\Gamma(\alpha)-\Gamma\left(\alpha,\frac{k}{\theta}\right)\right]+\lambda\mathrm{E}(\omega_i)-\lambda\mathrm{E}\left(e^{\omega_i}\right)-\lambda\ln\left(\frac{1}{\lambda}\right)-\ln\Gamma(\lambda)\right\}\right\}$$

and

$$\mathrm{E}(l_2) = \sum_{i=1}^{n}\left\{\sum_{j=1}^{n_i}\delta_{ij}\left\{-\frac{\mathrm{E}\left(-g(t_{ij})\right)+k}{\theta}+(\alpha-1)\,\mathrm{E}\left(\ln\left(-g(t_{ij})+k\right)\right)+\mathrm{E}\left(\ln\left(g'(t_{ij})\right)\right)\right\}\right.$$
$$\left.+(1-\delta_{ij})\,\mathrm{E}\left(\ln\left[\Gamma(\alpha)-\Gamma\left(\alpha,\frac{-g(t_{ij})+k}{\theta}\right)\right]\right)\right\}.$$

Since $\mathrm{E}\left(\ln\left(x\right)\right) \leq \ln\left(\mathrm{E}(x)\right)$ by Jensen's Inequality, the following relationship also holds.

$$\mathrm{E}(l_1)+\mathrm{E}(l_2) \leq \mathrm{E}(l_1)+\sum_{i=1}^{n}\sum_{j=1}^{n_i}\delta_{ij}\left\{-\frac{-(e^{-x_{ij}^T\beta}t_{ij})\mathrm{E}(e^{-\omega_i})+k}{\theta}\right.$$
$$+(\alpha-1)\ln\left(-\left(e^{-x_{ij}^T\beta}t_{ij}\right)\mathrm{E}(e^{-\omega_i})+k\right)-x_{ij}^T\beta-\mathrm{E}(\omega_i)\Big\}$$
$$\left.+(1-\delta_{ij})\ln\left(\Gamma(\alpha)-\mathrm{E}\left(\Gamma\left(\alpha,\frac{-(e^{-x_{ij}^T\beta}t_{ij})(e^{-\omega_i})+k}{\theta}\right)\right)\right)\right.$$

The conditional expectations of ω_i and its functions do not have closed form representations. Based on the conditional distribution of ω_i, which is proportional to (6.5), we sample ω_i using a Metropolis-Hastings algorithm. The Metropolis algorithm

was originally introduced by Metropolis, Rosenbluth, Rosenbluth, Teller, and Teller in 1953 and was generalized by Hastings in 1970. The algorithm generates samples from a distribution from which direct sampling is difficult. We use the generated samples to approximate the needed expectations.

M-Step:

The M-step is used to maximize $\mathrm{E}(l(\Theta)) = \mathrm{E}(l_1) + \mathrm{E}(l_2)$with respect to the unknown parameters by making use of the quantities found in the E-step. We make use of the fully specified survival, hazard and density functions.

Estimation procedure

Step 1: Choose initial values $\alpha^0, \beta^0, \theta^0, k^0, \lambda^0$.

Step 2: Sample ω_i from the posterior distribution and compute $\mathrm{E}(\omega_i), \mathrm{E}(e^{\omega_i})$ and $\mathrm{E}(e^{-\omega_i})$.

Step 3: Estimate new parameter values $\alpha^*, \boldsymbol{\beta}^*, \theta^*, k^*, \lambda^*$ by maximizing the likelihood function.

Step 4: Update the values of $\alpha, \boldsymbol{\beta}, \theta, k, \lambda$ and repeat steps 2 and 3 until the estimates converge.

6.4 Application to Pediatric Nephrotic Syndrome

In this section, we revisit the pediatric nephrotic syndrome data from New Jersey. We use the concept of individual frailty to assess the impact that unmeasured or non-measurable covariates at the individual level may have on the time to initial remission. The algorithm discussed in the previous section is used to find the maximum likelihood estimates and standard errors given in Table 6.1. We compare estimates from the frailty model to those of the AFT RSTG model without frailty given in

Table 6.1: RSTG AFT frailty model for pediatric nephrotic syndrome patients in New Jersey.

Distribution	Parameter	MLE(se)
RSTG	β_1	-0.042 (0.010)
	β_2	0.152 (0.036)
	β_3	1.089 (0.165)
	λ	8.192 (0.046)

Table 5.2. Age remains significant in the presence of individual frailty, but the effect is less pronounced. The decrease in time to first remission per one year increase in age is only 4%(95% CI :$2\% - 6\%$) after controlling for hematuria status and creatinine level and accounting for individual frailty. Hematuria status and creatinine level become significant in the presence of individual frailty. After controlling for age and creatinine level and accounting for individual frailty, the time to first remission for pediatric NS patients who exhibit hematuria at diagnosis is 16% longer (95% CI: $9\% - 25\%$) than that of patients who do not exhibit hematuria at diagnosis. After controlling for age and hematuria status and accounting for individual frailty, the time to first remission for the pediatric NS patient increases by approximately 12%(95% CI:$8\% - 15\%$) for each 0.1 mg/dL increase in creatinine level at diagnosis. The frailty variance is significantly larger than 0 in this model and suggests the presence of significant unobserved heterogeneity at the individual level ($\lambda = 8.192, SE = 0.046$).

The effect of hematuria status, which is the most influential of the covariates assessed on time to initial remission, supports findings of the original study.

6.5 The Correlated Frailty Model for Bivariate Data with Application to Diabetic Retinopathy

In this section, we evaluate a study on diabetic retinopathy in both juvenile and adult patients. The Diabetic Retinopathy Study (DRS) was begun by the National Eye

Institute in 1972 to study the effectiveness of laser photocoagulation in delaying the onset of blindness in patients with diabetic retinopathy. Patients were followed over several years for the occurrence of blindness in the left and right eye. The total study size was 1742. We consider the 50% sample ($N = 197$) of both juvenile and adult high risk patients as defined by DRS criteria that was first analyzed in 1989 by Huster et al. For each eye, the event of interest was the time from initiation of treatment to the time when visual acuity dropped below 5/200 for two consecutive visits (defined as "blindness"). Thus there is a built-in lag time of approximately 6 months (visits were every 3 months). Survival times in this dataset are the actual time to blindness in months, minus the minimum possible time to event (6.5 months). Censoring was caused by death, dropout, or end of the study.

Covariates considered are the laser photocoagulation treatment (0 = xenon, 1 = argon), age (in years), and diabetes diagnosis type (0=juvenile, 1=adult), with follow-up time given in months. The censoring indicator of each patient (0=censored, 1=blind) is also recorded. The distribution of times is negatively skewed with skewness coefficient of -0.33. For illustrative purposes, we consider in the first phase of analysis the times to blindness of the treated eye for each patient. We then account for the association between eyes of each patient by considering times to blindness in both the treated and untreated eye using a correlated frailty model.

6.5.1 The RSTG AFT model without frailty

We analyze the DRS data under the accelerated failure time framework with different distributional assumptions. We consider the covariates laser type, diabetic diagnosis type and the interaction. The AFT model is given by

$$\log(T_i) = \beta_0 + \beta_1 laser_i + \beta_2 diagnosis_i + \beta_3 laser_i \times diagnosis_i + \sigma\epsilon_i, \qquad (6.6)$$

where T_i represents the time to blindness in the treated eye of the i^{th} patient, $i = 1, ..., n$. Plots of the deviance residuals for the RSTG AFT model indicate no outliers, but

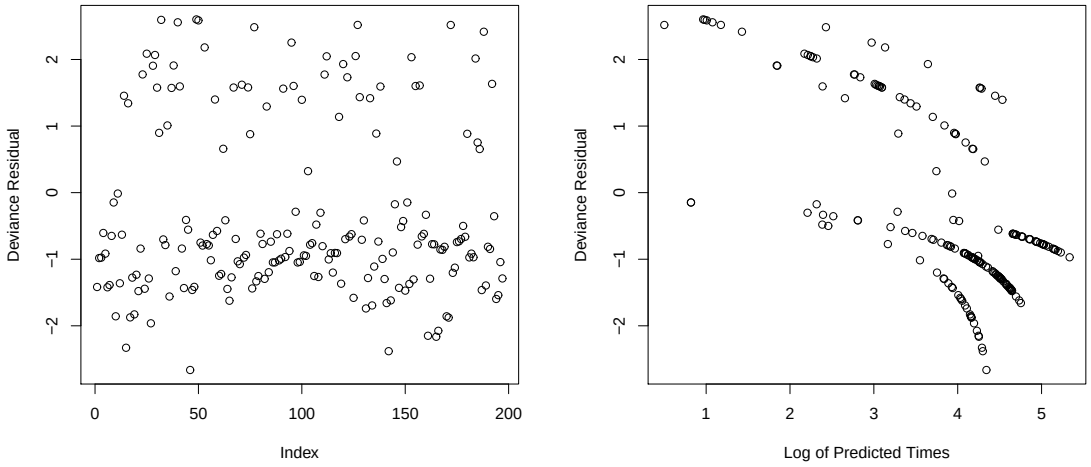

Figure 6.1: Deviance residual plots for the RSTG AFT model for the 1972 Diabetic Retinopathy Study.

Table 6.2: RSTG AFT model for the 1972 Diabetic Retinopathy Study.

Distribution	Parameter	MLE(se)
RSTG	β_1	-0.003 (0.422)
	β_2	0.581 (0.517)
	β_3	-0.919 (0.641)

suggest slight inadequacies in the model fit as the model tends to predict slightly longer times to blindness than are observed (Figure 6.1). This may be due to the 73% censoring rate present in the data set. The RSTG distributional assumption is best of the compared distributions (Appendix F).

No significant relationships between the time to blindness in the treated eye and laser type, diabetic diagnosis type or their interaction were detected by the model (Table 6.2).

6.5.2 The Correlated Gamma Frailty Model using the RSTG distribution

A fundamental consideration in choosing a strategy for the analysis of paired survival data is whether the correlation within a pair is a nuisance parameter or a parameter

of intrinsic scientific interest (Huster et al., 1989). In this section, we analyze the DRS data under the correlated gamma frailty model, a model introduced by Yashin, Vaupel, and Iachine (1995), using the RSTG survival function. We consider the time to blindness in both the treated and the untreated eye while adjusting for the correlation between the left and right eye of each patient. A primary goal of the DRS study was to assess the effectiveness of the photocoagulation treatment. The DRS Research Group (1976) reported that either photocoagulation treatment as carried out in this study was beneficial in reducing severe visual loss over a two-year period. This data set was later analyzed by Huster et al. (1989) and by Terry M Therneau and Grambsch (2000) under the proportional hazards framework with semiparametric Gaussian and gamma frailty models. The DRS data was also analyzed using a shared inverse Gaussian frailty model by Hanagal and Sharma (2013). Refer to the respective articles for a complete discussion of the results.

Following the methods discussed by Wienke (2010) for bivariate data, we note the following representation of the correlated gamma frailty model:

$$S(t_1, t_2) = \frac{S_1(t_1)^{1-\rho} S_2(t_2)^{1-\rho}}{(S_1(t_1)^{-\sigma^2} + S_2(t_2)^{-\sigma^2} - 1)^{\frac{\rho}{\sigma^2}}} \tag{6.7}$$

where ρ represents the frailty correlation between the left and right eye of each patient, $S_j(t), j = 1, 2$ represents the survival functions for the left eye and right eyes, and σ^2 represents the variance of the frailty random variable that is assumed to be the same for both eyes of each respective patient. Partial derivatives of the bivariate survival function can be found in the appendix of the Wienke text (2010). We use the simplifications $S(t) = S_1(t) = S_2(t)$ and use the RSTG baseline survival function. We obtain parameter estimates using the method of maximum likelihood (Table 6.3).

The results show a significant positive frailty correlation between the left and right eyes of each patient. A significant positive correlation was also noted in other studies (Sahu and Dey, 2000; Hanagal and Sharma, 2013). The inclusion of correlated frailty in the model produces a change in the significance of the laser type and the

Table 6.3: The RSTG distribution in a correlated gamma frailty model for the 1972 Diabetic Retinopathy Study.

Parameter	MLE(se)
β_1	1.288 (0.386)
β_2	0.071 (0.381)
β_3	1.419 (0.306)
ρ	0.403 (0.131)
σ^2	1.405 (1.310)

interaction of laser type and diagnosis type. The risk of blindness is significantly higher for those treated with the argon treatment as opposed to the xenon treatment ($\beta_1 = 1.288, SE = 0.386$) after controlling for diagnosis type. The risk is even more pronounced when patients are diagnosed with adult diabetes as opposed to juvenile diabetes ($\beta_3 = 1.419, SE = 0.306$) . The analysis by Hanagal and Sharma (2013) and Terry M Therneau and Grambsch (2000) both found a higher risk of blindness in the argon group and in the adult group individually, but the results were not statistically significant and did not include an interactive effect. Later analysts of diabetic retinopathy reported that treatment with xenon was associated with a higher rate of complications than argon laser and thus recommended argon laser photocoagulation treatment (Paulus and Blumenkranz, 2013).

6.6 Discussion

The model developed in Section 6.2 can serve as an individual frailty model in which the frailty represents either an unobserved individual effect or the event dependence of repeated events for each individual. In either case, the individual serves as a clustering unit. The developed model can also represent a shared frailty model in which the frailty is related to a specific characteristic that is shared by a group of individuals.

The current study makes use the gamma frailty distribution, but other flexible frailty distributions will be considered in future research. We employ the bootstrap method for variance estimation of the parameters in the individual frailty model. Further research is needed to investigate non-simulation based variance estimation techniques. Parameter estimates were obtained using the method of maximum likelihood, but the maximum likelihood approach may encounter difficulty when used in the frailty model (Hanagal et al., 2015). Bayesian approaches for parameter estimation are also viable options for frailty models (Ibrahim, Chen, and Sinha, 2005; Santos and Achcar, 2010).

Chapter 7

Summary and Future Goals

7.1 Summary

Negatively skewed survival data arise in medical research when data cluster near an upper limit. Simulation studies and comparisons using existing data sets show that the RSTG distribution performs better than the exponential, generalized gamma, generalized F, lognormal, log-logistic, Rayleigh, Gompertz and Weibull distributions when modeling negatively skewed data. The RSTG distribution also performs well when compared to the Gompertz-sinh family and the logistic-sinh family, which are two current alternative distributions designed to handle negatively skewed survival data. The RSTG distribution performs well as a baseline distribution for the general AFT model, for the AFT frailty model with gamma frailty, and for the correlated gamma frailty model. The brief example presented on the RSTG distribution used in a correlated gamma frailty model and applied to the 1972 DRS data had findings similar to those reported from previous analysis of the data.

Pediatric nephrotic syndrome is a rare disease syndrome that commonly has a relapsing course. Patients diagnosed with FRNS or SDNS, who previous research suggests experience a longer time to first remission, pose a greater challenge to healthcare providers in terms of disease management. Using a meta-analysis of worldwide studies, we detected a significant relationship between atopy and pediatric nephrotic syndrome. A study of South Carolina pediatric NS data was conducted with the RSTG AFT model to determine possible age or seasonal predictors of time to

first hospital visit after an NS diagnosis, but no significant predictors were found. Our study of the New Jersey pediatric NS data, conducted with the RSTG AFT model with individual frailty, shows that higher creatinine levels at diagnosis, presence of hematuria at diagnosis, and a younger age at diagnosis are indicative of a longer time to first remission for pediatric NS patients.

7.2 Future Goals

The majority of data on pediatric NS originates in areas outside of the U.S. We will provide a descriptive analysis of pediatric NS in South Carolina using the HSSC CDW and continue to search for predictive factors of the syndrome.

Patterns of relapse are a point of interest for pediatric NS patients. Previous studies suggest that relapse in the first year is a powerful independent predictor of subsequent relapse regardless of the duration of the illness (Atsushi Takeda, Takimoto, Mizusawa, and Simoda, 2001). More recent studies have concluded that a decrease in time from remission of the syndrome to first recurrence of symptoms predicts for a frequently relapsing course (Sureshkumar et al., 2014). Relapse of NS is almost universally defined as having proteinuria for three consecutive days after initial remission. Proteinuria can be detected outside of a clinical setting with the use of prescribed reagent strips for urinalysis, or in the clinical setting with urinalysis or blood tests. Another factor that may indicate relapse of the nephrotic syndrome is the presence of edema. While there is some discrepancy in the literature involving the nature of the edema, it is a condition that will most likely present itself if the nephrotic syndrome is left untreated and may be the first indication in the absence of a urinalysis that a relapse of the syndrome has occurred. Regardless of the method of detection, the relapse will have most likely occurred before an official clinical diagnosis was made, but within a time frame that was close to the time of diagnosis. Hence, we expect that the time-to-relapse originates

from a finite interval over which the distribution of times is negatively skewed. The RSTG distribution may be suitable to model these interval-censored event times.

Frailty models will be studied further to identify predictive factors of relapse or remission of the syndrome. A shared frailty model that uses sites such as the Robert Wood Johnson medical site as a clustering unit will be investigated. A repeated measures frailty model will be used on time-to-relapse data.

Additional properties of the RSTG distribution will be investigated. Alternative parameter estimation techniques for the RSTG distribution in the AFT model are other goals for continued study of this distribution. Also, the efficacy of the RSTG distribution in the Cox proportional hazards model and the other regression models will be investigated. The RSTG distribution can be used in any application as a distribution of choice for modeling event times arising from a negatively skewed distribution. An R package will be created to house the RSTG distribution and its associated functions.

Bibliography

Aalen, O. O. (1994). Effects of frailty in survival analysis. *Statistical Methods in Medical Research*, *3*(3), 227–243.

Abbring, J. H. & Van Den Berg, G. J. (2007). The unobserved heterogeneity distribution in duration analysis. *Biometrika*, *94*(1), 87–99.

Abdel-Hafez, M., Shimada, M., Lee, P. Y., Johnson, R. J., & Garin, E. H. (2009). Idiopathic nephrotic syndrome and atopy: is there a common link? *American Journal of Kidney Diseases*, *54*(5), 945–953.

About RWJMS. (2016). Retrieved May 11, 2016, from http://rwjms.rutgers.edu/about_rwjms/index.html

Al Salloum, A. A., Muthanna, A., Bassrawi, R., Al Shehab, A. A., Al Ibrahim, A., Islam, M. Z., & Al Hasan, K. (2012). Long-term outcome of the difficult nephrotic syndrome in children. *Saudi Journal of Kidney Diseases and Transplantation : an official publication of the Saudi Center for Organ Transplantation*, *23*(5), 965–72. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/22982908

American Academy of Pediatrics. (2015). Cholesterol levels in children and adolescents. Retrieved May 6, 2016, from https://www.healthychildren.org/English/healthy-living/nutrition/Pages/Cholesterol-Levels-in-Children-and-Adolescents.aspx

Andersen, P. K., Klein, J. P., Knudsen, K. M., & y Palacios, R. T. (1997). Estimation of variance in Cox's regression model with shared gamma frailties. *Biometrics*, 1475–1484.

Andersen, R. F., Thrane, N., Noergaard, K., Rytter, L., Jespersen, B., & Rittig, S. (2010). Early age at debut is a predictor of steroid-dependent and frequent relapsing nephrotic syndrome. *Pediatric Nephrology*, *25*(7), 1299–1304.

Anderson, D., Burnham, K., & White, G. (1998). Comparison of Akaike information criterion and consistent Akaike information criterion for model selection and statistical inference from capture-recapture studies. *Journal of Applied Statistics*, *25*(2), 263–282.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, 171–178.

Balakrishnan, N. & Peng, Y. (2006). Generalized gamma frailty model. *Statistics in Medicine*, *25*(16), 2797–2816.

Barratt, T., Osofsky, S., Bercowsky, A., Soothill, J., & Kay, R. (1975). Cyclophosphamide treatment in steroid-sensitive nephrotic syndrome of childhood. *The Lancet*, *305*(7898), 55–58.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2011). *Introduction to meta-analysis*. John Wiley and Sons.

Box-Steffensmeier, J. M. & De Boef, S. (2006). Repeated events survival models: the conditional frailty model. *Statistics in Medicine*, *25*(20), 3518.

Buckley, J. & James, I. (1979). Linear regression with censored data. *Biometrika*, *66*(3), 429–436.

Burnham, K. P. & Anderson, D. R. (2002). *Model selection and multimodel inference : a practical information-theoretic approach* (Second). New York: Springer.

Chen, P., Zhang, J., & Zhang, R. (2013). Estimation of the accelerated failure time frailty model under generalized gamma frailty. *Computational Statistics & Data Analysis*, *62*, 171–180.

Cheung, W., Wei Cl Fau - Seah, C.-C., Seah Cc Fau - Jordan, S. C., Jordan Sc Fau - Yap, H.-K., & Yap, H. K. (2004). Atopy, serum IgE, and interleukin-13 in steroid-responsive nephrotic syndrome. *Pediatric Nephrology*, *19*(6).

Childhood Nephrotic Syndrome. (2016). Retrieved May 10, 2016, from https://www.kidney.org/atoz/content/childns

Clayton, D. G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, *65*(1), 141–151.

Clinical Data Warehouse. (2016). Retrieved May 10, 2016, from http://www.healthsciencessc.org/cdw.asp

Collett, D. (2015). *Modelling survival data in medical research* (Third). CRC press.

Constantinescu, A. R., Shah, H. B., Foote, E. F., & Weiss, L. S. (2000). Predicting first-year relapses in children with nephrotic syndrome. *Pediatrics*, *105*(3), 492–495.

Cooray, K. (2005). Analyzing lifetime data with long-tailed skewed distribution: the logistic-sinh family. *Statistical Modelling*, *5*(4), 343–358.

Cooray, K. & Ananda, M. M. (2010). Analyzing survival data with highly negatively skewed distribution: the Gompertz-sinh family. *Journal of Applied Statistics*, *37*(1), 1–11.

Cordeiro, G. M., Ortega, E. M., & Silva, G. O. (2011). The exponentiated generalized gamma distribution with application to lifetime data. *Journal of Statistical Computation and Simulation*, *81*(7), 827–842.

Cox, D. R. [D. R.]. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, *34*(2), 187–220. Retrieved from http://www.jstor.org/stable/2985181

Cox, D. R. [David Roxbee] & Oakes, D. (1984). *Analysis of survival data*. CRC Press.

Crowther, M. J. & Lambert, P. C. (2013). Simulating biologically plausible complex survival data. *Statistics in Medicine*, *32*(23), 4118–34.

CVS Pharmacy. (2016). Prednisone side effects. Patient prescription leaflet.

de Mouzon-Cambon, A., Bouissou, F., Dutau, G., Barthe, P., Parra, M., Sevin, A., & Ohayon, E. (1981). HLA-DR7 in children with idiopathic nephrotic syndrome: correlation with atopy. *Tissue Antigens*, *17*(5), 518–524.

de Pascoa, M. A., Ortega, E. M., & Cordeiro, G. M. (2011). The Kumaraswamy generalized gamma distribution with application in survival analysis. *Statistical Methodology*, *8*(5), 411–433.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (methodological)*, 1–38.

Diabetic Retinopathy Study Research Group and others. (1976). Preliminary report on effects of photocoagulation therapy. *American Journal of Ophthalmology*, *81*(4), 383–396.

Dietz, W. H. & Stern, L. (Eds.). (2011). *Nutrition: what every parent needs to know*. American Academy of Pediatrics. Retrieved from http://ebooks.aappublications.org/content/9781581106312/9781581106312

Duchateau, L. & Janssen, P. (2007). *The frailty model*. Springer Science & Business Media.

Durkan, A. M., Hodson, E. M., Willis, N. S., & Craig, J. C. (2001). Immunosuppressive agents in childhood nephrotic syndrome: a meta-analysis of randomized controlled trials. *Kidney International*, *59*(5), 1919–1927. Retrieved from http://dx.doi.org/10.1046/j.1523-1755.2001.0590051919.x

Feudtner, C., Levin, J. E., Srivastava, R., Goodman, D. M., Slonim, A. D., Sharma, V., ... Hall, M. (2009). How well can hospital readmission be predicted in a cohort of hospitalized children? A retrospective, multicenter study. *Pediatrics*, *123*(1), 286–293.

Finkelstein, D. M. & Wolfe, R. A. (1985). A semiparametric model for regression analysis of interval-censored failure time data. *Biometrics*, *41*(4), 933–945

Fomina, S., Pavlenko, T., Englund, E., & Bagdasarova, I. (2011). Clinical course of steroid sensitive nephrotic syndrome in children: outcome and outlook. *Open Pediatric Medicine Journal*, *5*, 18–28.

Fujinaga, S., Hirano, D., & Nishizaki, N. (2011). Early identification of steroid dependency in Japanese children with steroid-sensitive nephrotic syndrome undergoing short-term initial steroid therapy. *Pediatric Nephrology*, *26*(3), 485–486

Gadegbeku, C. A., Gipson, D. S., Holzman, L. B., Ojo, A. O., Song, P. X., Barisoni, L., ... Kretzler, M. (2013). Design of the Nephrotic Syndrome Study Network (NEPTUNE) to evaluate primary glomerular nephropathy by a multidisciplinary approach. *Kidney International*, *83*(4), 749–56.

Gay, J. C., Hain, P. D., Grantham, J. A., & Saville, B. R. (2011). Epidemiology of 15-day readmissions to a children's hospital. *Pediatrics*, peds–2010.

Gilchrist, W. (2000). *Statistical modelling with quantile functions*. CRC Press.

Gipson, D. S. [D. S.], Selewski, D. T., Massengill, S. F., Wickman, L., Messer, K. L., Herreshoff, E., ... DeWalt, D. A. (2013). Gaining the PROMIS perspective from children with nephrotic syndrome : a Midwest pediatric nephrology consortium study. *Health and Quality of Life Outcomes*, *11*, 30.

Gipson, D. S. [Debbie S.], Massengill, S. F., Yao, L., Nagaraj, S., Smoyer, W. E., Mahan, J. D., ... Greenbaum, L. A. (2009). Management of childhood onset nephrotic syndrome. *Pediatrics*, *124*(2), 747–757

Gompertz, B. (1825). On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. *Philosophical Transactions of the Royal Society of London*, 513–583.

Greenwood, M. & Yule, G. U. (1920). An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or of repeated accidents. *Journal of the Royal Statistical Society*, *83*(2), 255–279.

Gupta, R. D. & Gupta, R. C. (2008). Analyzing skewed data by power normal model. *Test*, *17*(1), 197–210.

Gutierrez, R. G. et al. (2002). Parametric frailty and shared frailty survival models. *Stata Journal*, *2*(1), 22–44.

Hanagal, D. D., Pandey, A., & Ganguly, A. (2015). Correlated gamma frailty models for bivariate survival data. *Communications in Statistics-Simulation and Computation*, *85*(15).

Hanagal, D. D. & Sharma, R. (2013). Analysis of diabetic retinopathy data using shared inverse Gaussian frailty model. *Model Assisted Statistics and Applications*, *8*(2), 103–119.

Hanagal, D. D. & Sharma, R. (2015). Comparison of frailty models for acute leukemia data under Gompertz baseline distribution. *Communications in Statistics-Theory and Methods*, *44*(7), 1338–1350.

Hannan, E. J. & Quinn, B. G. (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society. Series B (Methodological)*, 190–195.

Harambat, J., Godron, A., Ernould, S., Rigothier, C., Llanas, B., & Leroy, S. (2013). Prediction of steroid-sparing agent use in childhood idiopathic nephrotic syndrome. *Pediatric Nephrology*, *28*(4), 631–638

Hardy, R. J. & Thompson, S. G. (1998). Detecting and describing heterogeneity in meta-analysis. *Statistics in Medicine*, *17*(8), 841–56.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, *57*(1), 97–109.

Hens, N., Wienke, A., Aerts, M., & Molenberghs, G. (2009). The correlated and shared gamma frailty model for bivariate current status data: an illustration for cross-sectional serological data. *Statistics in Medicine*, *28*(22), 2785–800.

Hilmanto, D. (2007). Association of HLA class II and history of atopy and frequent relapse of childhood steroid-sensitive nephrotic syndrome. *Paediatrica Indonesiana*, *47*(2), 60–64.

Hougaard, P. (1986). A class of multivanate failure time distributions. *Biometrika*, *73*(3), 671–678.

Hougaard, P. (1986b). Survival models for heterogeneous populations derived from stable distributions. *Biometrika*, *73*(2), 387–396.

Hougaard, P. (1999). Fundamentals of survival data. *Biometrics*, *55*(1), 13–22.

Hougaard, P. (2012). *Analysis of multivariate survival data*. Springer Science & Business Media.

Hussain, N., Zello, J. A., Vasilevska-Ristovska, J., Banh, T., Patel, V., Patel, P., ... Parekh, R. (2013). The rationale and design of Insight into Nephrotic Syndrome: Investigating Genes, Health and Therapeutics (INSIGHT): a prospective cohort study of childhood nephrotic syndrome. *BMC Nephrology*, *14*(1), 25. Retrieved from http://www.biomedcentral.com/1471-2369/14/25

Huster, W. J., Brookmeyer, R., & Self, S. G. (1989). Modelling paired survival data with covariates. *Biometrics*, 145–156.

Hutton, J. & Monaghan, P. (2002). Choice of parametric accelerated life and proportional hazards models for survival data: asymptotic results. *Lifetime data analysis*, *8*(4), 375–393.

Ibrahim, J. G., Chen, M.-H., & Sinha, D. (2005). *Bayesian survival analysis*. Wiley Online Library.

Iijima, K., Hamahira, K., Tanaka, R., Kobayashi, A., Nozu, K., Nakamura, H., & Yoshikawa, N. (2002). Risk factors for cyclosporine-induced tubulointerstitial lesions in children with minimal change nephrotic syndrome. *Kidney International*, *61*(5), 1801–1805.

Ishikura, K., Yoshikawa, N., Nakazato, H., Sasaki, S., Iijima, K., Nakanishi, K., ... Honda, M. (2012). Two-year follow-up of a prospective clinical trial of cyclosporine for frequently relapsing nephrotic syndrome in children. *Clinical Journal of the American Society of Nephrology*, *7*(10), 1576–83.

Jahan, I., Hanif, M., Ali, M., Waliullah, S., & Mia, A. (2011). Relationship between serum IgE and frequent relapse idiopathic nephrotic syndrome. *Mymensingh Medical Journal*, *20*(3), 484–489.

Jin, Z., Lin, D., Wei, L., & Ying, Z. (2003). Rank-based inference for the accelerated failure time model. *Biometrika*, *90*(2), 341–353.

Johnson, N. L., Kotz, S., & Balakrishnan, N. (2002). *Continuous multivariate distributions, volume 1, models and applications*. New York: John Wiley & Sons.

Kalbfleisch, J. D. & Prentice, R. L. (2011). *The statistical analysis of failure time data*. John Wiley & Sons.

Keiding, N., Andersen, P. K., & Klein, J. P. (1997). The role of frailty models and accelerated failure time models in describing heterogeneity due to omitted covariates. *Statistics in Medicine*, *16*(2), 215–224.

Kelley, C. T. (1999). *Iterative methods for optimization*. Siam.

Kerlin, B. A., Haworth, K., & Smoyer, W. E. (2014). Venous thromboembolism in pediatric nephrotic syndrome. *Pediatric Nephrology*, *29*(6), 989–997.

Klein, J. P. [John P], Moeschberger, M., Li, Y., Wang, S., & Flournoy, N. (1992). Estimating random effects in the Framingham Heart Study. In *Survival analysis: state of the art* (pp. 99–120). Springer.

Klein, J. P. [John P.] & Moeschberger, M. L. (2003). *Survival analysis: techniques for censored and truncated data* (Second). Statistics for Biology and Health. New York: Springer-Verlag.

Kleinbaum, D. G. & Klein, M. (2006). *Survival analysis: a self-learning text*. Springer Science & Business Media.

Lambert, P., Collett, D., Kimber, A., & Johnson, R. (2004). Parametric accelerated failure time models with random effects and an application to kidney transplant survival. *Statistics in Medicine*, *23*(20), 3177–3192.

Larson, R., Hostetler, R., Edwards, B., & Heyd, D. (2002). *Calculus with analytic geometry*. Houghton Mifflin Company, Boston, NY.

Latta, K., von Schnakenburg, C., & Ehrich, J. H. (2001). A meta-analysis of cytotoxic treatment for frequently relapsing nephrotic syndrome in children. *Pediatric Nephrology*, *16*(3), 271–282.

Lee, E. T. & Wang, J. (2003). *Statistical methods for survival data analysis* (Third). Hoboken, New Jersey: John Wiley & Sons.

Letavernier, B., Letavernier, E., Leroy, S., Baudet-Bonneville, V., Bensman, A., & Ulinski, T. (2008). Prediction of high-degree steroid dependency in pediatric idiopathic nephrotic syndrome. *Pediatric Nephrology*, *23*(12), 2221–2226

Lin, C.-Y., Lee, B.-H., Lin, C.-C., & Chen, W.-P. (1990). A study of the relationship between childhood nephrotic syndrome and allergic diseases. *CHEST Journal*, *97*(6), 1408–1411.

Liu, L., Wolfe, R. A., & Huang, X. (2004). Shared frailty models for recurrent events and a terminal event. *Biometrics*, *60*(3), 747–756.

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

Lombel, R. M., Gipson, D. S., & Hodson, E. M. (2013). Treatment of steroid-sensitive nephrotic syndrome: new guidelines from KDIGO. *Pediatric Nephrology*, *28*(3), 415–26.

Lucena, S. E., Silva, A. H. A., & Cordeiro, G. M. (2015). The transmuted generalized gamma distribution: properties and application. *Journal of Data Science*, *13*(1), 193–212.

Maiti, S. S. & Dey, M. (2012). Tilted normal distribution and its survival properties. *Journal of Data Science*, *10*(2), 225–240.

McGilchrist, C. & Aisbett, C. (1991). Regression with frailty in survival analysis. *Biometrics*, 461–466.

Meadow, S. R. & Sarsfield, J. K. (1981). Steroid-responsive and nephrotic syndrome and allergy: clinical studies. *Archives of Disease in Childhood*, *56*(7), 509–516. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pmcentrez&artid=1627348

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, *21*(6), 1087–1092.

Mishra, O. P., Abhinay, A., Mishra, R. N., Prasad, R., & Pohl, M. (2013). Can we predict relapses in children with idiopathic steroid-sensitive nephrotic syndrome? *Journal of Tropical Pediatrics*, *59*(5), 343–9.

Nakanishi, K., Iijima, K., Ishikura, K., Hataya, H., Nakazato, H., Sasaki, S., ... Yoshikawa, N. (2013). Two-year outcome of the ISKDC regimen and frequent-relapsing risk in children with idiopathic nephrotic syndrome. *Clinical Journal of the American Society of Nephrology*, *8*(5), 756–62.

National Institute of Health. (2014). *Glomerular diseases*. Publication No.14-4358. National Kidney and Urologic Diseases Clearinghouse.

National Institute of Health. (2016). Nephrotic syndrome: Rare Diseases Clinical Research Network (RDCRN). Retrieved May 7, 2016, from https://www.rarediseasesnetwork.org/cms/neptune/Learn-More/Disorder-Definitions

Noer, M. S. (2005). Predictors of relapse in steroid sensitive nephrotic syndrome. *Southeast Asian Journal of Tropical Medicine and Public Health*, *36*(5), 1313.

Oakes, D. (1989). Bivariate survival models induced by frailties. *Journal of the American Statistical Association*, *84*(406), 487–493.

Oakes, D. (1992). Frailty models for multiple event times. *Survival analysis: State of the art*, *4*, 371–379.

Ofungwu, J. (2014). *Statistical applications for environmental analysis and risk assessment*. John Wiley & Sons.

Orbe, J., Ferreira, E., & Núñez-Antón, V. (2002). Comparing proportional hazards and accelerated failure time models for survival analysis. *Statistics in Medicine*, *21*(22), 3493–3510.

Pan, W. (2001). Using frailties in the accelerated failure time model. *Lifetime Data Analysis*, *7*(1), 55–64.

Pasini, A., Aceto, G., Ammenti, A., Ardissino, G., Azzolina, V., Bettinelli, A., ... On behalf of the NefroKid Study, G. (2015). Best practice guidelines for idiopathic nephrotic syndrome: recommendations versus reality. *Pediatric Nephrology (Berlin, Germany)*, *30*(1), 91–101

Patel, K., Kay, R., & Rowell, L. (2006). Comparing proportional hazards and accelerated failure time models: an application in influenza. *Pharmaceutical Statistics*, *5*(3), 213–224.

Paulus, Y. M. & Blumenkranz, M. S. (2013). Proliferative and nonproliferative diabetic retinopathy. *American Academy of Opthamology*. Retrieved from http://www.aao.org/munnerlyn-laser-surgery-center/laser-treatment-of-proliferative-nonproliferative-

Pediatric Nephrotic Syndrome. (2016). Retrieved May 10, 2016, from https://clinicaltrials.gov/ct2/results?term=pediatric+nephrotic+syndrome&cntry1=NA%3AUS

Pickles, A. & Crouchley, R. (1995). A comparison of frailty models for multivariate survival data. *Statistics in Medicine*, *14*(13), 1447–1461.

Pourhoseingholi, M. A., Hajizadeh, E., Moghimi Dehkordi, B., Safaee, A., Abadi, A., & Zali, M. R. (2007). Comparing Cox regression and parametric models for survival of patients with gastric carcinoma. *Asian Pacific Journal of Cancer Prevention*, *8*(3), 412.

Pradhan, B. & Kundu, D. (2011). Bayes estimation and prediction of the two-parameter gamma distribution. *Journal of Statistical Computation and Simulation*, *81*(9), 1187–1198.

Prentice, R. L. (1978). Linear rank tests with right censored data. *Biometrika*, *65*(1), 167–179.

R Core Team. (2016). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. Retrieved from https://www.R-project.org/

Rebien, W., Müller-Wiefel, D., Wahn, U., & Schärer, K. (1981). IgE mediated hypersensitivity in children with idiopathic nephrotic syndrome. *The International Journal of Pediatric Nephrology*, *2*(1), 23–28.

Research. (2016). Retrieved May 10, 2016, from http://www.healthsciencessc.org/research.asp

Richardson, M. A. (2012). The many faces of minimal change nephrotic syndrome: an overview and case study. *Nephrology Nursing Journal*, *39*(5), 365–74, quiz 375.

Ritov, Y. (1990). Estimation in a linear regression model with censored data. *The Annals of Statistics*, 303–328.

Robertson, H. T. & Allison, D. B. (2012). A novel generalized normal distribution for human longevity and other negatively skewed data. *PloS one*, *7*(5), e37025.

Robins, J. (1992). Estimation of the time-dependent accelerated failure time model in the presence of confounding factors. *Biometrika*, *79*(2), 321–334.

Roy, R. R., Islam, M. R., Matin, A., Khan, R., Muinuddi, G., Rahman, M. H., & Hossain, M. M. (2012). Relationship of childhood idiopathic nephrotic syndrome with asthma, hypertension, complement C 3, urinalysis. *Bangladesh Journal of Child Health*, *35*(1), 11–15.

Rubio, F. J. & Steel, M. F. (2012). On the Marshall–Olkin transformation as a skewing mechanism. *Computational Statistics and Data Analysis*, *56*(7), 2251–2257.

Sahu, S. K. & Dey, D. K. (2000). A comparison of frailty and other models for bivariate survival data. *Lifetime Data Analysis*, *6*(3), 207–228.

Saleem, M. A. (2013). New developments in steroid-resistant nephrotic syndrome. *Pediatric Nephrology*, *28*(5), 699–709.

Salsano, M. E., Luisa, G., Ilaria, L., Paola, P., Mario, G., & Giuliana, L. (2007). Atopy in childhood idiopathic nephrotic syndrome. *Acta Paediatrica*, *96*(4), 561–566

Santos, C. & Achcar, J. (2010). A bayesian analysis for multivariate survival data in the presence of covariates. *Journal of Statistical Theory and Applications*, *9*, 233–253.

Sarker, M., Islam, M., Saad, T., Shoma, F., Sharmin, L., Khan, H., ... Salimullah, A., et al. (2012). Risk factor for relapse in childhood nephrotic syndrome - A hospital based retrospective study. *Faridpur Medical College Journal*, *7*(1), 18–22.

Sinha, A., Bhatia, D., Gulati, A., Rawat, M., Dinda, A. K., Hari, P., & Bagga, A. (2015). Efficacy and safety of rituximab in children with difficult-to-treat nephrotic syndrome. *Nephrology Dialysis Transplantation*, *30*(1), 96–106

Soyka, L. F. (1967). Treatment of the nephrotic syndrome in childhood: use of an alternate-day prednisone regimen. *American Journal of Diseases of Children*, *113*(6), 693–701

Sprent, P. & Smeeton, N. C. (2007). *Applied nonparametric statistical methods* (Fourth). Boca Raton, FL: CRC press.

Stacy, E. W. (1962). A generalization of the gamma distribution. *The Annals of Mathematical Statistics*, 1187–1192.

Sureshkumar, P., Hodson, E. M., Willis, N. S., Barzi, F., & Craig, J. C. (2014). Predictors of remission and relapse in idiopathic nephrotic syndrome: a prospective cohort study. *Pediatric Nephrology*, *29*(6), 1039–1046.

Swindell, W. R. (2009). Accelerated failure time models provide a useful statistical framework for aging research. *Experimental Gerontology*, *44*(3), 190–200.

Tain, Y.-L., Chen, T.-Y., & Yang, K. D. (2003). Implication of serum IgE in childhood nephrotic syndrome. *Pediatric Nephrology*, *18*(12), 1211–1215.

Takeda, A. [A.], Matsutani, H., Niimura, F., & Ohgushi, H. (1996). Risk factors for relapse in childhood nephrotic syndrome. *Pediatric Nephrology*, *10*(6), 740–1.

Takeda, A. [Atsushi], Takimoto, H., Mizusawa, Y., & Simoda, M. (2001). Prediction of subsequent relapse in children with steroid-sensitive nephrotic syndrome. *Pediatric Nephrology*, *16*(11), 888–893.

Tarshish, P., Tobin, J. N., Bernstein, J., & Edelmann, C. M. (1997). Prognostic significance of the early course of minimal change nephrotic syndrome: report of the International Study of Kidney Disease in Children. *Journal of the American Society of Nephrology*, *8*(5), 769–76. Retrieved from http://jasn.asnjournals.org/content/8/5/769.abstract

Tenbrock, K., Schubert, A., Stapenhorst, L., Kemper, M., Gellermann, J., Timmermann, K., ... Michalk, D. (2002). Type I IgE receptor, interleukin 4 receptor and interleukin 13 polymorphisms in children with nephrotic syndrome. *Clinical Science*, *102*(5), 507–512.

Thatcher, A. R. (1999). The long-term pattern of adult mortality and the highest attained age. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *162*(1), 5–43.

Therneau, T. M. [Terry M] & Grambsch, P. M. (2000). *Modeling survival data: extending the cox model*. Springer Science & Business Media.

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

Therneau, T. M. [Terry M], Grambsch, P. M., & Pankratz, V. S. (2003). Penalized survival models and frailty. *Journal of Computational and Graphical Statistics*, *12*(1), 156–175.

Thomson, P., Stokes, C., Barratt, T., Turner, M., & Soothill, J. (1976). HLA antigens and atopic features in steroid-responsive nephrotic syndrome of childhood. *The Lancet*, *308*(7989), 765–768.

Toyabe, S.-i., Nakamizo, M., Uchiyama, M., & Akazawa, K. (2005). Circannual variation in the onset and relapse of steroid-sensitive nephrotic syndrome. *Pediatric Nephrology*, *20*(4), 470–473.

Trompeter, R. S., Barratt, T. M., Kay, R., Turner, M. W., & Soothill, J. F. (1980). HLA, atopy, and cyclophosphamide in steroid-responsive childhood nephrotic syndrome. *Kidney International*, *17*(1), 113–117.

Tsiatis, A. A. (1990). Estimating regression parameters using linear rank tests for censored data. *The Annals of Statistics*, 354–372.

Tullus, K. & Marks, S. D. (2013). Indications for use and safety of rituximab in childhood renal diseases. *Pediatric Nephrology*, *28*(7), 1001–1009.

U.S. National Library of Medicine. (2016). Medline plus: comprehensive metabolic panel. Retrieved May 6, 2016, from https://www.nlm.nih.gov/medlineplus/ency/article/003468.htm

Vaupel, J. W., Manton, K. G., & Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, *16*(3), 439–454.

Vivarelli, M., Moscaritolo, E., Tsalkidis, A., Massella, L., & Emma, F. (2010). Time for initial response to steroids is a major prognostic factor in idiopathic nephrotic syndrome. *The Journal of Pediatrics*, *156*(6), 965–971

Wang, Y.-p., Liu, A.-m., Dai, Y.-w., Yang, C., & Tang, H.-f. (2005). The treatment of relapsing primary nephrotic syndrome in children. *Journal of Zhejiang University. Science. B*, *6*(7), 682–685

Wei, L. (1992). The accelerated failure time model: a useful alternative to the Cox regression model in survival analysis. *Statistics in Medicine*, *11*(14-15), 1871–1879.

Wienke, A. (2010). *Frailty models in survival analysis*. CRC Press.

Yap, H., Han, E. J., Heng, C.-K., & Gong, W.-K. (2001). Risk factors for steroid dependency in children with idiopathic nephrotic syndrome. *Pediatric Nephrology*, *16*(12), 1049–1052.

Yap, H., Yip, W., Lee, B., Ho, T., Teo, J., Aw, S., & Tay, J. (1983). The incidence of atopy in steroid-responsive nephrotic syndrome: clinical and immunological parameters. *Annals of Allergy*, *51*(6), 590–594.

Yashin, A. I., Vaupel, J. W., & Iachine, I. A. (1995). Correlated individual frailty: an advantageous approach to survival analysis of bivariate data. *Mathematical Population Studies*, *5*(2), 145–159.

Youn, Y. S. & Lim, H. H. (2012). The clinical characteristics of steroid responsive nephrotic syndrome of children according to the serum immunoglobulin E levels and cytokines. *Yonsei Medical Journal*, *53*(4), 715–722. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pmcentrez&artid=3381495

Zhang, J. & Peng, Y. (2007). An alternative estimation method for the accelerated failure time frailty model. *Computational Statistics and Data Analysis*, *51*(9), 4413–4423.

Appendix A

Nephrotic Syndrome Definitions

Steroid Sensitive Nephrotic Syndrome - complete remission within 4 weeks after initiation of standard corticosteroid therapy

Minimal Change Disease - term used to classify the most common biopsy characterization of the disease

Steroid Responsive Nephrotic Syndrome - also referred to as Steroid Sensitive Nephrotic Syndrome

Steroid Resistant Nephrotic Syndrome - persistent edema and proteinuria (failure to achieve complete remission) after 8 weeks of standard corticosteroid therapy.

Steroid Dependent Nephrotic Syndrome - two consecutive relapses during corticosteroid therapy, or within 14 days of ceasing therapy

Frequently Relapsing Nephrotic Syndrome - two or more relapses within 6 months of initial response to corticosteroid therapy, or four or more relapses in any 12-month period.

Infrequently Relapsing Nephrotic Syndrome - one relapse within 6 months of initial response to corticosteroid therapy, or one to three relapses in any 12-month period.

Appendix B

Studies Used in Meta-Analysis

1. Saleem, M. A. (2013). New developments in steroid-resistant nephrotic syndrome. *Pediatric Nephrology*, *28*(5), 699–709.

2. National Institute of Health. (2016). Nephrotic syndrome: Rare Diseases Clinical Research Network (RDCRN).. Retrieved May 7, 2016, from https://www.rarediseasesnetwork.org/cms/neptune/Learn-More/Disorder-Definitions .

3. Constantinescu, A. R., Shah, H. B., Foote, E. F., & Weiss, L. S. (2000). Predicting first-year relapses in children with nephrotic syndrome. *Pediatrics*, *105*(3), 492–495.

4. Meadow, S. R. & Sarsfield, J. K. (1981). Steroid-responsive and nephrotic syndrome and allergy: clinical studies. *Archives of Disease in Childhood*, *56*(7), 509–516. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pmcentrez&artid=1627348 .

5. Salsano, M. E., Luisa, G., Ilaria, L., Paola, P., Mario, G., & Giuliana, L. (2007). Atopy in childhood idiopathic nephrotic syndrome. *Acta Paediatrica*, *96*(4), 561–566.

6. Cheung, W., Wei Cl Fau - Seah, C.-C., Seah Cc Fau - Jordan, S. C., Jordan Sc Fau - Yap, H.-K., & Yap, H. K. (2004). Atopy, serum IgE, and interleukin-13 in steroid-responsive nephrotic syndrome. *Pediatric Nephrology*, *19*(6).

7. Sarker, M., Islam, M., Saad, T., Shoma, F., Sharmin, L., Khan, H., ... Salimullah, A., et al. (2012). Risk factor for relapse in childhood nephrotic syndrome - A hospital based retrospective study. *Faridpur Medical College Journal*, *7*(1), 18–22.

8. Yap, H., Han, E. J., Heng, C.-K., & Gong, W.-K. (2001). Risk factors for steroid dependency in children with idiopathic nephrotic syndrome. *Pediatric Nephrology*, *16*(12), 1049–1052.

9. Al Salloum, A. A., Muthanna, A., Bassrawi, R., Al Shehab, A. A., Al Ibrahim, A., Islam, M. Z., & Al Hasan, K. (2012). Long-term outcome of the difficult nephrotic syndrome in children. *Saudi Journal of Kidney Diseases and Transplantation : an official publication of the Saudi Center for Organ Transplantation*, *23*(5), 965–72. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/22982908 .

10. Hilmanto, D. (2007). Association of HLA class II and history of atopy and frequent relapse of childhood steroid-sensitive nephrotic syndrome. *Paediatrica Indonesiana*, *47*(2), 60–64.

11. Trompeter, R. S., Barratt, T. M., Kay, R., Turner, M. W., & Soothill, J. F. (1980). HLA, atopy, and cyclophosphamide in steroid-responsive childhood nephrotic syndrome. *Kidney International*, *17*(1), 113–117.

12. Yap, H., Yip, W., Lee, B., Ho, T., Teo, J., Aw, S., & Tay, J. (1983). The incidence of atopy in steroid-responsive nephrotic syndrome: clinical and immunological parameters. *Annals of Allergy*, *51*(6), 590–594.

13. Youn, Y. S. & Lim, H. H. (2012). The clinical characteristics of steroid responsive nephrotic syndrome of children according to the serum immunoglobulin E levels and cytokines. *Yonsei Medical Journal*, *53*(4), 715–722. Retrieved from http:

//www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pmcentrez&artid=3381495 .

14. Lin, C.-Y., Lee, B.-H., Lin, C.-C., & Chen, W.-P. (1990). A study of the relationship between childhood nephrotic syndrome and allergic diseases. *CHEST Journal*, *97*(6), 1408–1411.

15. Tain, Y.-L., Chen, T.-Y., & Yang, K. D. (2003). Implication of serum IgE in childhood nephrotic syndrome. *Pediatric Nephrology*, *18*(12), 1211–1215.

16. Tenbrock, K., Schubert, A., Stapenhorst, L., Kemper, M., Gellermann, J., Timmermann, K., ... & Michalk, D. (2002). Type I IgE receptor, interleukin 4 receptor and interleukin 13 polymorphisms in children with nephrotic syndrome. *Clinical Science*, *102*(5), 507–512.

17. Mishra, O. P., Abhinay, A., Mishra, R. N., Prasad, R., & Pohl, M. (2013). Can we predict relapses in children with idiopathic steroid-sensitive nephrotic syndrome? *Journal of Tropical Pediatrics*, *59*(5), 343–9.

18. Toyabe, S.-i., Nakamizo, M., Uchiyama, M., & Akazawa, K. (2005). Circannual variation in the onset and relapse of steroid-sensitive nephrotic syndrome. *Pediatric Nephrology*, *20*(4), 470–473.

19. Roy, R. R., Islam, M. R., Matin, A., Khan, R., Muinuddi, G., Rahman, M. H., & Hossain, M. M. (2012). Relationship of childhood idiopathic nephrotic syndrome with asthma, hypertension, complement C 3, urinalysis. *Bangladesh Journal of Child Health*, *35*(1), 11–15.

20. de Mouzon-Cambon, A., Bouissou, F., Dutau, G., Barthe, P., Parra, M., Sevin, A., & Ohayon, E. (1981). HLA-DR7 in children with idiopathic nephrotic syndrome: correlation with atopy. *Tissue Antigens*, *17*(5), 518–524.

21. Thomson, P., Stokes, C., Barratt, T., Turner, M., & Soothill, J. (1976). HLA antigens and atopic features in steroid-responsive nephrotic syndrome of childhood. *The Lancet*, *308*(7989), 765–768.

22. Rebien, W., Müller-Wiefel, D., Wahn, U., & Schärer, K. (1981). IgE mediated hypersensitivity in children with idiopathic nephrotic syndrome. *The International Journal of Pediatric Nephrology*, *2*(1), 23–28.

23. Jahan, I., Hanif, M., Ali, M., Waliullah, S., & Mia, A. (2011). Relationship between serum IgE and frequent relapse idiopathic nephrotic syndrome. *Mymensingh Medical Journal*, *20*(3), 484–489.

24. National Institute of Health. (2014). *Glomerular diseases*. Publication No.14-4358. National Kidney and Urologic Diseases Clearinghouse.

25. Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2011). *Introduction to meta-analysis*. John Wiley and Sons.

26. Barratt, T., Osofsky, S., Bercowsky, A., Soothill, J., & Kay, R. (1975). Cyclophosphamide treatment in steroid-sensitive nephrotic syndrome of childhood. *The Lancet*, *305*(7898), 55–58.

27. Hardy, R. J. & Thompson, S. G. (1998). Detecting and describing heterogeneity in meta-analysis. *Statistics in Medicine*, *17*(8), 841–56.

Appendix C

Derivations of the RSTG Distribution Functions

C.1 Probability Density Function

The two-parameter gamma distribution is given by

$$f(t|\alpha,\theta) = \frac{1}{\Gamma(\alpha)\theta^{\alpha}} e^{-\frac{t}{\theta}} t^{\alpha-1}, 0 < t < \infty,$$

where $\alpha > 0$ represents the shape parameter and $\theta > 0$ represents the scale parameter. Reflecting the two-parameter gamma distribution about the y-axis and shifting it $k > 0$, the shift parameter, units to the right gives a probability density function of

$$f_1(t|\alpha,\theta,k) = \frac{1}{\Gamma(\alpha)\theta^{\alpha}} e^{-\frac{(-t+k)}{\theta}} (-t+k)^{\alpha-1}, -\infty < t < k, \alpha, \theta > 0.$$

The cumulative distribution function of this reflected, shifted gamma distribution is

$$\begin{aligned}
F_1(t|\alpha,\theta,k) &= \int_{-\infty}^{t} \frac{1}{\Gamma(\alpha)\theta^{\alpha}} e^{-\frac{(-x+k)}{\theta}} (-x+k)^{\alpha-1} \mathrm{d}x \\
&= \int_{-\infty}^{t} \frac{(-x+k)^{\alpha-1}}{\Gamma(\alpha)\theta^{\alpha}} e^{-\frac{(-x+k)}{\theta}} \mathrm{d}x \\
&= \frac{1}{\Gamma(\alpha)} \int_{-\infty}^{t} \frac{(-x+k)^{\alpha-1}}{\theta^{\alpha-1}\cdot\theta} e^{-\frac{(-x+k)}{\theta}} \mathrm{d}x \\
&= \frac{1}{\Gamma(\alpha)} \int_{-\infty}^{t} \left[\frac{1}{\theta}(-x+k)\right]^{\alpha-1} \cdot e^{-\frac{(-x+k)}{\theta}} \cdot \frac{1}{\theta} \mathrm{d}x
\end{aligned}$$

Let $u = \frac{1}{\theta}(-x+k)$. Then $\mathrm{d}u = -\frac{1}{\theta}\mathrm{d}x$. Changing the limits of integration from those in terms of x to those in terms of u we have

$$\begin{aligned} F_1(t|\alpha,\theta,k) &= \frac{1}{\Gamma(\alpha)} \int_{\infty}^{\frac{-t+k}{\theta}} u^{\alpha-1}e^{-u}(-\mathrm{d}u) \\ &= \frac{1}{\Gamma(\alpha)} \int_{\frac{-t+k}{\theta}}^{\infty} u^{\alpha-1}e^{-u}\mathrm{d}u \\ &= \frac{1}{\Gamma(\alpha)} \left[\Gamma\left(\alpha, \frac{-t+k}{\theta}\right)\right] \end{aligned}$$

for $t < k$ where $\Gamma(a) = \int_0^{\infty} t^{a-1}e^{-t}\,\mathrm{d}t$ is the gamma function and $\Gamma(a,b) = \int_b^{\infty} t^{a-1}e^{-t}\,\mathrm{d}t$ represents the upper incomplete gamma function.

We now truncate the reflected, shifted gamma distribution on the left at 0, restricting the interval for t to $[0,k]$. To achieve a valid probability density function, we divide by the area that remains after the truncation, $F_1(k) - F_1(0)$.

The probability density function of the reflected-shifted-truncated-gamma (RSTG) distribution, then, is

$$\begin{aligned} f^*(t|\alpha,\theta,k) &= \frac{1}{F_1(k)-F_1(0)}\left(\frac{1}{\Gamma(\alpha)\theta^{\alpha}}\right)e^{-\frac{(-t+k)}{\theta}}(-t+k)^{\alpha-1} \\ &= \frac{1}{\frac{\Gamma(\alpha)}{\Gamma(\alpha)} - \frac{\Gamma(\alpha,\frac{k}{\theta})}{\Gamma(\alpha)}}\left(\frac{1}{\Gamma(\alpha)\theta^{\alpha}}\right)e^{-\frac{(-t+k)}{\theta}}(-t+k)^{\alpha-1} \\ &= \frac{1}{\frac{\Gamma(\alpha)-\Gamma(\alpha,\frac{k}{\theta})}{\Gamma(\alpha)}}\left(\frac{1}{\Gamma(\alpha)\theta^{\alpha}}\right)e^{-\frac{(-t+k)}{\theta}}(-t+k)^{\alpha-1} \\ &= \frac{\Gamma(\alpha)}{\Gamma(\alpha)-\Gamma(\alpha,\frac{k}{\theta})}\left(\frac{1}{\Gamma(\alpha)\theta^{\alpha}}\right)e^{-\frac{(-t+k)}{\theta}}(-t+k)^{\alpha-1} \\ &= \frac{1}{\theta^{\alpha}\left(\Gamma(\alpha)-\Gamma(\alpha,\frac{k}{\theta})\right)}e^{-\frac{(-t+k)}{\theta}}(-t+k)^{\alpha-1} \quad \alpha>0,\ \theta>0,\ 0\le t<k \end{aligned}$$

C.2 Cumulative Distribution Function

The CDF of the RSTG distribution is given by

$$\begin{aligned}
F^*(t|\alpha,\theta,k) &= \frac{1}{\theta^\alpha\left(\Gamma(\alpha)-\Gamma(\alpha,\frac{k}{\theta})\right)}\int_0^t e^{-\frac{(-x+k)}{\theta}}(-x+k)^{\alpha-1}\,\mathrm{d}x \\
&= \frac{1}{\left(\Gamma(\alpha)-\Gamma(\alpha,\frac{k}{\theta})\right)}\int_0^t \frac{(-x+k)^{\alpha-1}}{\theta^\alpha}e^{-\frac{(-x+k)}{\theta}}\,\mathrm{d}x \\
&= \frac{1}{\left(\Gamma(\alpha)-\Gamma(\alpha,\frac{k}{\theta})\right)}\int_0^t \frac{(-x+k)^{\alpha-1}}{\theta^{\alpha-1}\cdot\theta}e^{-\frac{(-x+k)}{\theta}}\,\mathrm{d}x \\
&= \frac{1}{\left(\Gamma(\alpha)-\Gamma(\alpha,\frac{k}{\theta})\right)}\int_0^t \left[\frac{1}{\theta}(-x+k)\right]^{\alpha-1}\cdot e^{-\frac{(-x+k)}{\theta}}\cdot\frac{1}{\theta}\mathrm{d}x.
\end{aligned}$$

Let $u=\frac{1}{\theta}(-x+k)$. Then $\mathrm{d}u=-\frac{1}{\theta}\mathrm{d}x$. Changing the limits of integration from those in terms of x to those in terms of u we have

$$\begin{aligned}
F^*(t|\alpha,\theta,k) &= \frac{1}{\left(\Gamma(\alpha)-\Gamma(\alpha,\frac{k}{\theta})\right)}\int_{\frac{k}{\theta}}^{\frac{-t+k}{\theta}} u^{\alpha-1}\cdot e^{-u}\,(-\mathrm{d}u) \\
&= \frac{1}{\left(\Gamma(\alpha)-\Gamma(\alpha,\frac{k}{\theta})\right)}\int_{\frac{-t+k}{\theta}}^{\frac{k}{\theta}} u^{\alpha-1}\cdot e^{-u}\,\mathrm{d}u \\
&= \frac{1}{\left(\Gamma(\alpha)-\Gamma(\alpha,\frac{k}{\theta})\right)}\left[\int_0^{\infty} u^{\alpha-1}\cdot e^{-u}\mathrm{d}u-\int_0^{\frac{-t+k}{\theta}} u^{\alpha-1}\cdot e^{-u}\mathrm{d}u\right. \\
&\qquad\qquad \left.-\int_{\frac{k}{\theta}}^{\infty} u^{\alpha-1}\cdot e^{-u}\,\mathrm{d}u\right].
\end{aligned}$$

The lower incomplete gamma function is given by $\gamma(a,b)=\int_0^b t^{a-1}e^{-t}\,\mathrm{d}t$. Thus,

$$\begin{aligned}
F^*(t|\alpha,\theta,k) &= \frac{1}{\left(\Gamma(\alpha)-\Gamma(\alpha,\frac{k}{\theta})\right)}\left[\Gamma(\alpha)-\gamma\left(\alpha,\frac{-t+k}{\theta}\right)-\Gamma\left(\alpha,\frac{k}{\theta}\right)\right] \\
&= \frac{1}{\left(\Gamma(\alpha)-\Gamma(\alpha,\frac{k}{\theta})\right)}\left[\Gamma(\alpha)-\left[\Gamma(\alpha)-\Gamma\left(\alpha,\frac{-t+k}{\theta}\right)\right]-\Gamma\left(\alpha,\frac{k}{\theta}\right)\right] \\
&= \frac{\Gamma\left(\alpha,\frac{-t+k}{\theta}\right)-\Gamma\left(\alpha,\frac{k}{\theta}\right)}{\Gamma(\alpha)-\Gamma\left(\alpha,\frac{k}{\theta}\right)}.
\end{aligned}$$

Appendix D

Hessian Matrix of the RSTG Distribution

The Hessian matrix can be written as

$$H(\alpha,\theta,k)=\begin{pmatrix}\dfrac{\partial^2 l}{\partial\alpha^2} & \dfrac{\partial^2 l}{\partial\alpha\partial\theta} & \dfrac{\partial^2 l}{\partial\alpha\partial k}\\ \dfrac{\partial^2 l}{\partial\theta\partial\alpha} & \dfrac{\partial^2 l}{\partial\theta^2} & \dfrac{\partial^2 l}{\partial\theta\partial k}\\ \dfrac{\partial^2 l}{\partial k\partial\alpha} & \dfrac{\partial^2 l}{\partial k\partial\theta} & \dfrac{\partial^2 l}{\partial k^2}\end{pmatrix}=\begin{pmatrix}H_{11}(\alpha,\theta,k) & H_{12}(\alpha,\theta,k) & H_{13}(\alpha,\theta,k)\\ H_{21}(\alpha,\theta,k) & H_{22}(\alpha,\theta,k) & H_{23}(\alpha,\theta,k)\\ H_{31}(\alpha,\theta,k) & H_{32}(\alpha,\theta,k) & H_{33}(\alpha,\theta,k)\end{pmatrix}.$$

with entries defined as

$$H_{11}(\alpha,\theta,k)=-n\left\{\frac{\left(\Gamma(\alpha)-\Gamma\left(\alpha,\frac{k}{\theta}\right)\right)\left[\Gamma''(\alpha)-\Gamma''\left(\alpha,\frac{k}{\theta}\right)\right]-\left(\Gamma'(\alpha)-\Gamma'\left(\alpha,\frac{k}{\theta}\right)\right)^2}{\left(\Gamma(\alpha)-\Gamma\left(\alpha,\frac{k}{\theta}\right)\right)^2}\right\}$$

$$H_{12}(\alpha,\theta,k)=-n\left\{\frac{1}{\theta}+\frac{\left(\Gamma(\alpha)-\Gamma\left(\alpha,\frac{k}{\theta}\right)\right)\left[-\frac{1}{\theta}e^{-\frac{k}{\theta}}\ln\left(\frac{k}{\theta}\right)\left(\frac{k}{\theta}\right)^{\alpha}\right]}{\left(\Gamma(\alpha)-\Gamma\left(\alpha,\frac{k}{\theta}\right)\right)^2}\right\}$$
$$-n\left\{\frac{\left(\frac{1}{\theta}\left(\frac{k}{\theta}\right)^{\alpha}e^{-\frac{k}{\theta}}\right)\left[\Gamma'(\alpha)-\Gamma'\left(\alpha,\frac{k}{\theta}\right)\right]}{\left(\Gamma(\alpha)-\Gamma\left(\alpha,\frac{k}{\theta}\right)\right)^2}\right\}$$

$$H_{13}(\alpha,\theta,k)=-n\left\{\frac{\left(\Gamma(\alpha)-\Gamma\left(\alpha,\frac{k}{\theta}\right)\right)\left[\frac{1}{k}\ln\left(\frac{k}{\theta}\right)\left(\frac{k}{\theta}\right)^{\alpha}e^{-\frac{k}{\theta}}\right]}{\left(\Gamma(\alpha)-\Gamma\left(\alpha,\frac{k}{\theta}\right)\right)^2}\right\}$$
$$-n\left\{\frac{\left[\Gamma'(\alpha)-\Gamma'\left(\alpha,\frac{k}{\theta}\right)\right]\left[\frac{1}{k}\left(\frac{k}{\theta}\right)^{\alpha}e^{-\frac{k}{\theta}}\right]}{\left(\Gamma(\alpha)-\Gamma\left(\alpha,\frac{k}{\theta}\right)\right)^2}\right\}$$

$$H_{22}(\alpha,\theta,k) = -n\left\{ -\frac{\alpha}{\theta^2} + \frac{\frac{k^\alpha}{\theta^{\alpha+2}}e^{-\frac{k}{\theta}}\left(-\frac{k}{\theta^\alpha}+\alpha+1\right)\left(\Gamma(\alpha)-\Gamma\left(\alpha,\frac{k}{\theta}\right)\right)+\frac{k^{2\alpha}}{\theta^{2(\theta+1)}}e^{-\frac{2k}{\theta}}}{\left(\Gamma(\alpha)-\Gamma\left(\alpha,\frac{k}{\theta}\right)\right)^2}\right\}$$
$$-\frac{2\sum_{i=1}^{n}(-t_i+k)}{\theta^3}$$

$$H_{23}(\alpha,\theta,k) = -n\left\{\frac{\left(\Gamma(\alpha)-\Gamma\left(\alpha,\frac{k}{\theta}\right)\right)\left[\frac{1}{k\theta^2}\left(k\left(\frac{k}{\theta}\right)^\alpha e^{-\frac{k}{\theta}}-\alpha\theta\left(\frac{k}{\theta}\right)^\alpha e^{-\frac{k}{\theta}}\right)\right]}{\left(\Gamma(\alpha)-\Gamma\left(\alpha,\frac{k}{\theta}\right)\right)^2}\right\}$$
$$-n\left\{\frac{\frac{-1}{\theta}\left(\frac{k}{\theta}\right)^\alpha e^{-\frac{k}{\theta}}\frac{1}{k}\left(\frac{k}{\theta}\right)^\alpha e^{-\frac{k}{\theta}}}{\left(\Gamma(\alpha)-\Gamma\left(\alpha,\frac{k}{\theta}\right)\right)^2}\right\}+\frac{n}{\theta^2}$$

$$H_{33}(\alpha,\theta,k) = \frac{\left(\Gamma(\alpha)-\Gamma\left(\alpha,\frac{k}{\theta}\right)\right)\left[\frac{n}{k}\left(\frac{k}{\theta}\right)^{\alpha-1}\left(\frac{1}{\theta}\right)e^{-\frac{k}{\theta}}\left(\frac{k}{\theta}+\alpha-1\right)\right]-\frac{n}{k^2}\left(\frac{k}{\theta}\right)^{2\alpha}e^{-\frac{2k}{\theta}}}{\left(\Gamma(\alpha)-\Gamma\left(\alpha,\frac{k}{\theta}\right)\right)^2}$$

where $\Gamma'(\alpha) = \frac{d\Gamma(\alpha)}{d\alpha} = \psi(\alpha)\Gamma(\alpha)$, $\Gamma''(\alpha) = \frac{d^2\Gamma(\alpha)}{d^2\alpha} = \Gamma(\alpha)\psi(\alpha,1)+\psi^2(\alpha)\Gamma(\alpha)$, $\Gamma'(\alpha,\frac{k}{\theta}) = \frac{d\Gamma(\alpha,\frac{k}{\theta})}{d\alpha} = \int_{\frac{k}{\theta}}^{\infty}\ln(y)y^{\alpha-1}e^{-y}\mathrm{d}y$, and $\Gamma''(\alpha,\frac{k}{\theta}) = \frac{d^2\Gamma(\alpha,\frac{k}{\theta})}{d^2\alpha} = \int_{\frac{k}{\theta}}^{\infty}\ln^2(y)y^{\alpha-1}e^{-y}\mathrm{d}y$. Also, $\psi(\alpha) = \frac{d\Gamma(\ln(\alpha))}{d\alpha}$ and $\psi(\alpha,1) = \frac{d\psi(\alpha)}{d\alpha}$.

We assume the existence of Clairut's theorem on the equality of mixed partial derivatives.

Theorem 1 (Clairut's Theorem)**.** *If f is a function of x and y such that $\frac{\partial^2 f}{\partial x\partial y}$ and $\frac{\partial^2 f}{\partial y\partial x}$ are continuous on an open disc R, then, for every (x,y) in R,*

$$\frac{\partial^2 f}{\partial x\partial y} = \frac{\partial^2 f}{\partial y\partial x}.$$

This theorem applies to a function f of three or more variables as long as the second partial derivatives are continuous (Larson, Hostetler, Edwards, and Heyd, 2002).

Appendix E

Common Continuous Distributions

Distribution	**Density Function** $f(t)$	**Parameters**
Exponential	$\lambda e^{-\lambda t}$	$\lambda > 0$
Generalized F	$\frac{\delta \left(\frac{s_1}{s_2}\right)^{s_1} e^{s_1 w}}{\sigma t \left(1 + s_1 \frac{e^w}{s_2}\right)^{s_1+s_2} B(s_1, s_2)}$	$\mu, Q, P \in \mathbb{R}, \sigma > 0$ $s_1 = 2(Q^2 + 2P + Q\delta)^{-1}$ $s_2 = 2(Q^2 + 2P - Q\delta)^{-1}$ $\delta = (Q^2 + 2P)^{\frac{1}{2}}$ $w = \frac{(\log t - \mu)\delta}{\sigma}$
Generalized Gamma	$\frac{\lvert Q\rvert (Q^{-2})^{Q^{-2}}}{\sigma t \Gamma(Q^{-2})} exp[Q^{-2}(Qw - e^{Qw})]$	$\sigma > 0, \mu, Q \in \mathbb{R}$ $w = \log(Q^2\gamma)/Q$ and $\gamma \sim$ Gamma$(Q^{-2}, 1)$ for $t = exp(\mu + \sigma w)$
Gompertz	$be^{at}e^{\frac{b}{a}(1-e^{bt})}$	$a \in \mathbb{R}$ $b > 0$
Log-logistic	$\frac{\frac{\alpha}{\beta}\left(\frac{t}{\beta}\right)^{\alpha-1}}{\left(1 + \left(\frac{t}{\beta}\right)^{\alpha}\right)^2}$	$\alpha, \beta > 0$
Lognormal	$\frac{1}{\sigma t\sqrt{2\pi}} e^{\frac{-(\log t - \mu)^2}{2\sigma^2}}$	$\mu \in \mathbb{R}$ $\sigma > 0$
Rayleigh	$\frac{t}{b^2} e^{-\frac{t^2}{2b^2}}$	$b > 0$
Weibull	$\lambda\gamma t^{\gamma-1} e^{-\lambda t^{\gamma}}$	$\lambda, \gamma > 0$

* $0 \leq t < \infty$, $B(\cdot)$ is the beta function, $\gamma(\cdot)$ is the incomplete gamma function

Appendix F

Simulation Study and Comparative Model Fits of the RSTG AFT Model

F.1 Simulation Study

For each of 1000 repetitions, we simulate 200 negatively skewed survival times, a continuous covariate $x_1 \sim N(2, 0.12)$ and a categorical covariate $x_2 \sim Bin(2, 0.5)$. Following an assumption used by Crowther and Lambert (2013) to simulate biologically plausible data, we assume administrative censoring to achieve a less than 20% censoring rate.

We fit an accelerated failure time model to the simulated data using the RSTG, exponential, generalized gamma, generalized F, lognormal, log-logistic, Rayleigh, Gompertz and Weibull baseline distributions. We compare them in Table F.1 using the information theoretic criteria as defined in Chapter 4. The average covariate parameter estimate, average standard error of the parameter estimate, and average AIC values of the simulations are reported in Table F.1.

The covariate parameter estimates of the RSTG distribution appear to be more stable than the parameter estimates given by the comparison models. We note that while the Gompertz model is commonly used for left skewed distributions, the AIC value of the RSTG model is more than 10 units lower than that of the Gompertz model and of other compared models. The RSTG AFT model is the superior model based on the simulated right-censored data.

Table F.1: Model fitting results for simulated data.

Distribution	Par	MLE(se)	AIC	AICC	HQIC	CAIC
Exponential	β_1	0.0442 (1.9220)	648.25	648.29	649.78	651.86
	β_2	0.0009 (0.9467)				
Generalized	β_1	0.0526 (0.1644)	507.45	507.59	508.98	521.87
F	β_2	-0.0019 (0.0273)				
Generalized	β_1	0.0224 (0.1429)	500.51	500.62	502.04	511.33
Gamma	β_2	-0.0039 (0.0232)				
Gompertz	β_1	-0.0066 (0.1529)	476.60	476.68	478.13	483.81
	β_2	0.0016 (0.0257)				
Log-logistic	β_1	-0.0214 (0.3290)	564.93	565.00	566.46	572.14
	β_2	0.0028 (0.0563)				
Lognormal	β_1	0.0384 (0.8458)	664.40	664.47	665.93	671.61
	β_2	-0.0037 (0.1445))				
Rayleigh	β_1	-0.0074 (0.9624)	570.39	570.43	571.92	573.00
	β_2	-0.0017 (0.4741)				
Weibull	β_1	-0.0052 (0.2540)	533.12	533.19	534.65	540.33
	β_2	0.0022 (0.0431)				
RSTG	β_1	-0.0230 (0.0242)	**465.37**	**465.44**	**466.89**	**472.58**
	β_2	0.0251 (0.0225)				

F.2 Comparative Model Fits for HSSC Data

The RSTG distributional assumption is more appropriate than the compared distributional assumptions for the HSSC data based on the AIC and information theoretic criteria (Table F.2). The AIC value of the RSTG distribution is more than 10 units lower than the compared distributions.

F.3 Comparative Model Fits for DRS Data

The RSTG AFT model provides the best fit for the DRS data based on the AIC and other information theoretic criteria of the compared distributions (Table F.3). The AIC value of the RSTG distribution is more than 10 units lower than the compared distributions.

Table F.2: Model fitting results for the HSSC pediatric nephrotic syndrome data.

Distribution	Par	MLE(se)	AIC	AICC	HQIC	CAIC
Exponential	β_1	-0.037 (0.056)	206.75	207.27	208.03	219.80
	β_2	0.413 (0.530)				
	β_3	-0.381 (0.608)				
	β_4	-0.554 (0.785)				
Generalized F	β_1	0.071 (0.006)	188.16	189.05	189.44	209.05
	β_2	-0.079 (0.063)				
	β_3	-0.363 (0.088)				
	β_4	-0.474 (0.087)				
Generalized Gamma	β_1	0.070 (0.002)	184.50	185.26	185.78	202.78
	β_2	-0.085 (0.035)				
	β_3	-0.358 (0.038)				
	β_4	0.469 (0.048)				
Gompertz	β_1	-0.009 (0.017)	188.16	188.79	189.44	203.83
	β_2	0.184 (0.152)				
	β_3	-0.191 (0.216)				
	β_4	-0.198 (0.266)				
Log-logistic	β_1	-0.026 (0.023)	195.17	195.80	196.45	210.84
	β_2	0.146 (0.248)				
	β_3	-0.153 (0.247)				
	β_4	-0.302 (0.311)				
Lognormal	β_1	-0.037 (0.024)	194.82	195.45	196.10	210.49
	β_2	0.153 (0.255)				
	β_3	-0.070 (0.242)				
	β_4	-0.323 (0.319)				
Rayleigh	β_1	0.022 (0.027)	194.90	195.42	196.18	207.95
	β_2	-0.245 (0.265)				
	β_3	0.212 (0.305)				
	β_4	0.312 (0.392)				
Weibull	β_1	-0.017 (0.021)	193.45	193.97	194.73	206.50
	β_2	0.200 (0.684)				
	β_3	-0.170 (0.757)				
	β_4	-0.246 (0.808)				
RSTG	β_1	0.046 (0.060)	**169.84**	**170.73**	**171.12**	**190.73**
	β_2	3.153 (2.820)				
	β_3	1.384 (1.499)				
	β_4	3.138 (4.420)				

Table F.3: Model fitting results for treated eyes of patients in the 1972 Diabetic Retinopathy Study.

Distribution	Par	MLE(se)	AIC	AICC	HQIC	CAIC
Exponential	β_1	0.201 (0.002)	647.18	647.26	648.84	668.60
	β_2	-0.495 (-0.417)				
	β_3	-0.030 (0.579)				
Generalized	β_1	-0.348 (0.565)	640.20	640.33	641.86	674.47
F	β_2	0.802 (0.560)				
	β_3	-0.180 (0.831)				
Generalized	β_1	-0.538 (0.485)	637.42	637.54	639.08	667.41
Gamma	β_2	0.636 (0.517)				
	β_3	-0.029 (0.750)				
Gompertz	β_1	0.201 (0.335)	653.09	653.17	654.75	674.51
	β_2	-0.495 (0.417)				
	β_3	-0.030 (0.578)				
Log-logistic	β_1	0.223 (0.459)	647.22	647.30	648.88	668.64
	β_2	-0.694 (0.544)				
	β_3	0.015 (0.760)				
Lognormal	β_1	0.286 (0.476)	643.52	643.62	645.18	669.22
	β_2	-0.721 (0.550)				
	β_3	0.028 (0.775)				
Rayleigh	β_1	-0.123 (0.168)	727.53	727.61	729.19	748.95
	β_2	0.235 (0.208)				
	β_3	0.073 (0.289)				
Weibull	β_1	0.244 (0.425)	649.03	649.11	650.69	670.45
	β_2	-0.620 (0.532)				
	β_3	-0.009 (0.732)				
RSTG	β_1	- 0.003 (0.422)	**599.96**	**600.08**	**601.62**	**629.94**
	β_2	0.581 (0.517)				
	β_3	-0.919 (0.641)				