Multivariate Survival Analysis (I): Shared Frailty Approaches to Reliability and Dependence Modeling

Zhanshan (Sam) Ma sam@cs.uidaho.edu
Axel W. Krings krings@cs.uidaho.edu
Computer Science Department
University of Idaho
Moscow, ID 83844, USA

Abstract—The latest advances in survival analysis have been centered on multivariate systems. Multivariate survival analysis has two major categories of models: one is multi-state modeling; the other is shared frailty modeling. Multi-state models, although formulated differently in both fields, have been extensively studied in reliability analysis in the context of Markov chain analysis. In contrast, shared frailty modeling seems little known in reliability analysis and computer science. In this article, we focus exclusively on shared frailty modeling. Shared frailty refers to the often-unobserved factors or risks responsible for the *common risks dependence* between multiple events. It is well recognized as the most effective modeling approach to address common risks dependence and, more recently, the *event-related dependence*. The only exclusion of dependence modeling for the frailty approach is the *common events* type, which is best addressed by multi-state modeling. We argue that shared frailty modeling not only is perfectly applicable for engineering reliability, but also is of significant potential in other fields of computer science, such as networking and software reliability and survivability, machine learning, and prognostics and health management (PHM).

INDEX TERMS: Multivariate Survival Analysis, Shared Frailty Model, Dependent Failure, Common Risks Failure, Common Events Failure, Event-Related Dependence, Network Survivability, Prognostic and Health Management (PHM), Software Reliability.

ABBREVIATION: MSA – Multivariate Survival Analysis
PHM – Prognostic and Health Management.

TABLE OF CONTENTS [1] [2]

[1] 1-4244-1488-1/08/$25.00 ©2008 IEEE
[2] IEEEAC paper #1650, Final Version Updated Dec. 27, 2007

1. INTRODUCTION

Survival analysis has a history of nearly four decades and has become the *de facto* standard in biomedical research. Survival analysis and reliability theory have the exact same mathematical models in their basic definitions. For example, survivor function and reliability have identical probability definitions, and both fields use the same term and definition formula for the hazard function. However, the two fields have diverged gradually over the years and relatively little interaction has happened between them. This lack of interaction is unfortunate in our opinion, since reliability and survival analysis essentially address the same mathematical problems—the study of time-to-event random variables in the more general abstraction. Whether the event is the failure of a device, the death or survival of a patient, or simply the occurrence of any time-to-event, the underlying mathematical model should be very similar.

This article is the third in a four part series, in which we review state-of-the-art research in *survival* (*univariate*) *analysis*, *competing risks analysis* and *multivariate survival analysis* as well as their applications to engineering reliability and computer science. The other three areas, univariate survival analysis, competing risks analysis, and multi-state modeling (Part-II of multivariate survival analysis) are addressed in Ma and Krings 2008a, 2008b & 2008c, respectively. The present paper intends to briefly introduce the essential concepts, models and methods of shared frailty modeling and present our suggestions and opinions on their potential applications to engineering and computer science fields.

As indicated by Hougaard (2000) in the very first monograph of MSA, which is certainly a significant milestone in the field, MSA is still in its infancy and its advancement has been closely related to the demands arising from biomedical and public health research. One should expect significant complications in MSA, compared with standard multivariate analysis. For example, unlike standard multivariate statistical analysis that can often be characterized by an $m \times n$ matrix as its basic data structure, there are at least five data structures (models or formats) in MSA. Data censoring and dependence between variables further complicate the analysis significantly.

Although the topic of MSA was raised as early as the 1980s (e.g., Cox 1984), major research activities in MSA were

started in the middle 1990s. Like survival analysis, the application papers of MSA spread over all major medical journals such as *JAMA*, *New England Journal of Medicine*, and *Surgery*. The methodologies of MSA are mostly published in *Biometrics*, *Biometrika*, *Statistics in Medicine*, *Statistical Methods in Medical Research*, *Biomedical Journal*, *Lifetime Data Analysis,* and several statistical journals. Houggard's (2000) *Analysis of Multivariate Survival Data* is still the only monograph in MSA if we treat competing risks analysis as a separate field. The following is a far from complete list of the review papers on multivariate survival analysis in last few years: Aalen 1994, Lin 1994, Harrell et al. 1996, Wei and Glidden 1997, Commenges 1999, Hougaard 1999a, 1999b, 2000, Oakes 2001, Monaco et al. 2005, and Escobar and Meeker 2006. While most of the afore-mentioned review papers are outside the field of engineering reliability, Lisnianski and Levitin's (2003) monograph reviews the multi-state models in reliability theory. The major monographs on univariate survival analysis also often contain a comprehensive chapter on multivariate survival analysis (Andersen et al. 1993, Fleming and Harrington 1991, Cox and Oakes. 1984, Kalbfleisch and Prentice 2002, Lawless 2003, Klein and Moeschberger 2003, and Ibrahim et al. 2005). Unlike univariate survival analysis, standard software packages, such as *SAS*® or *SPSS*®, have implemented very few procedures for performing MSA. The most common practice seems to be programming with *S*-plus or its open-source counterpart *R*.

One notion is that MSA is needed whenever the dependence between survival times cannot be ignored (Houggard 2001); this clearly demonstrates the central importance of the dependence issue in MSA. As Aven and Jensen (1999) summarized, dependence modeling is one of the most challenging works in reliability engineering. Therefore, the applicability of survival analysis in reliability engineering is beyond question. Although the application potential in reliability is often mentioned by survival analysis theoreticians, few actual applications of MSA have surfaced in engineering reliability.

To get a glimpse of the status of the applications of survival analysis in computer science and IEEE related engineering, we recently (in the July of 2007) conducted an online search of the IEEE Digital Library with the relevant keywords. Regarding univariate and competing risks analysis, we found approximately 40 papers with the keyword *survival analysis*, and 20 papers with *competing risks* in IEEE digital library, respectively. However, we did not get any paper in IEEE digital library with the keywords of either *multivariate survival analysis* or *frailty model*. Given this situation, the format of this paper is somewhat different from the previous two (Ma and Krings 2008a, b) where we had a dedicated section to review papers from IEEE digital library and selected papers from MMR-2004 (*International Conference on Mathematical Methods in Reliability*). Here, we introduce essential concepts in MSA and major frailty models (but skip multi-state models), based on the only monograph (Hougaard 2000), the latest review papers, and selected research papers in biomedical fields. We also try to present our opinions and suggestions on the potential applications of those concepts and models in computer science and engineering reliability.

2. DATA MODELS AND COVARIATES

2.1 Parallel and Longitudinal Data.

The basic notations, concepts and classifications follow Hougaard (2000). The so-called ***parallel data*** arise from research in which one follows several individuals or objects simultaneously. The lifetimes of a group of twins would be one of the simplest parallel data sets. An example in reliability would be the system consisting of parallel components. The data set then consists of a nxk matrix of T_{ij}, with $i = 1, 2, ..., n$, $j = 1, 2, ..., k$, where i denotes group, and j denotes an individual within the group. A corresponding set of failure indicators is needed, D_{ij}, $i = 1, 2, ..., n$, $j = 1, 2, ..., k$. As pointed out in Hougaard (2000), it is usually assumed that the groups are independent, but no restriction is imposed within a group. The number of *times* (k) in a group should be pre-specified. However, as some of the *times* correspond to failures and some to censoring, the number of events ($\sum_j D_{ij}$) is not specified. Certainly, the maximum number of events is k. Furthermore, there is no ordering requirement for the data, for example, $T_{i1} > T_{i2}$ is not required. The ordering restriction is only imposed on longitudinal data.

The meaning of *parallel* in MSA is obviously broader than that in engineering reliability, and it could be significantly different in both fields. Lifetimes of twins, failure times of similar organs (e.g., eyes, kidneys) are the most typical parallel data in biomedicine (Hougaard 2000). This type of data may be hard to get in biomedicine, but it is much more common in engineered systems. Perhaps this type of parallel data can be considered as the orthodox parallel model in reliability modeling.

Longitudinal data arise from the observations of stochastic processes (Hougaard 2000). One follows a collection of stochastic processes, one for each value of i, and observe the events or transitions over time. In biomedical research, there is one process for each person. Longitudinal data is also called life history data. Similar to parallel data, one assumes independence between groups. Thus, the times of transitions forms an increasing sequence of times. In the simplest cases, only the last time can be censored. All the others are events (times of transitions). The number of times (K_i) is random.

More precisely, in longitudinal data, one observes the process i over the interval $(0, C_i)$, where C_i is a fixed or random time. The times of events are in the form of T_{ij}, $i = 1, 2, ..., n$, $j = 1, 2, ... K_i$. Depending on the situations, one may let the times be the pure event times, using C_i as the

end of the observation. Alternatively, one may distinguish times as events times and a censoring time (C_i). The definition imposes the ordering requirement, that is, $T_{ij} \leq T_{i,j+1}$. The key difference from parallel data is then the imposition of the data ordering in longitudinal data (Hougaard 2000).

One type of longitudinal data is *recurrent events*. When a single individual experiences the same event multiple times, we can observe and record the recurrent events. Recurrent events may be widespread in software testing measurements, such as the discovery of 1, 2,...,*n* bugs or potentially infinite bugs and ultimately the observation is truncated due to the stop of the testing. The study of recurrent events has received very significant attention in the last few years. Peña (2006) presented an excellent review about this topic. As Peña (2006) indicated, the study of recurrent events data is complicated by several unique features, for example, the effects of accelerating number of events, covariates, dependence, and/or intervention after each occurrence, etc.

One interesting and extremely important phenomenon to note is that recurrent event is often ended with a terminal event, for example, the recurrence of cancer and the ultimate death of a patient. Liu et al. (2004) considered frailty proportional hazards models for the recurrent and terminal event processes. The dependence between recurrent events and terminal event is captured by *conditioning on* the shared frailty in both hazards functions.

Repeated measurements are often obtained from the carefully designed experiment. The individuals in the experiment are observed to experience the same type of event for a fixed number of times, and the times to the event are recorded. It is similar to the recurrent events in the sense that the same type of event is observed at longitudinal time points. Nevertheless, the data structure is parallel because the number of observation times is fixed. This data model assumes that the event an individual experiences does not harm the individual. Multi-state model does not work for this data model, since a given time point may correspond to several different points in calendar time. However, the shared frailty model is well suited for the repeated measurements data (Hougaard 2000). The repeated measurements model does not seem useful for hardware reliability analysis, since there is only one failure in hardware, which is often damaging or absorbing. However, there are two apparent application potentials in computer science. One is the software reliability analysis, for example, the repeated occurrences of same bug observed within a time period. Another application can be the analysis of the adaptive machine learning.

Different events data model: In previous data models, the events in observation are either the same or similar. There is certainly the need for more general multiple events data model. For example, a network server may have three states: healthy, infected (with virus) but still functional, and down states. One may be interested in one individual's states or the states of individuals in a population. One is particularly interested in how the various events influence the future course of an individual. For example, different events such as OS memory leaks, which become worse with time, potential hacker attacks, second infestation by virus, or even the increasing traffic load, could influence the course of the evolution of server states. Multi-state modeling is often the best option for representing this data model.

In biomedical research with animals, a well-known experiment is the survival-sacrifice experiment, where some animals are killed in order to assess the probability of experiencing a tumor at selected time points (Hougaard 2000). Similarly, in a large-scale Grid computing network, a sampling scheme can be used to estimate the states' distribution over time (Jafar and Krings et al. 2008).

Competing risks data refers to several potential or latent failure risks competing for the ultimate cause of the failure. Only one of the latent risks turns into the failure cause after the failure occurs, and the other latent risks lose their identities. What is observable is the failure time and corresponding failure cause, but either time or cause, or both, may be censored. Competing risks data is also called multiple decrement data in population demography and actuarial science (Crowder 2001, Hougaard 2000, Ma and Krings 2008b).

2.2 Covariates

Covariates are explanatory variables that influence failure or survival times. Some covariates are intrinsic such as treatments in drug trials; others are external factors, which may not be the focus of the study but, nevertheless, influence failures (Hougaard 2000). In software reliability analysis, covariates may be usage profiles, which are often beyond the control of the software developer but certainly influence whether or not a bug would be exposed. Operating system platforms and network environments, patch maintenance, and usage intensity might be characterized by the usage behaviors, which may stretch the design limits of software and be treated as external covariates. Other factors such as extreme file size and open file numbers, critical section access contentions, multiple and distributed access may be treated as intrinsic covariates since they should be considered by software designers.

In the parallel data model, covariates are represented as a *p*-dimension vector $[z_{ij}]_m$ for the *i,j-th* object and its component would be z_{ijm}. There are two special cases that are of particular significance: the common covariates and matched-pairs covariates. For parallel data, common covariates are the same for all members of the group. In the case of longitudinal data, they do not change as the process proceeds. In other words, common covariates depend on group (*i*) only. In contrast, matched pairs covariates are only dependent on *j*. They apply to parallel data, but are not applicable to the longitudinal data model, because one

cannot guarantee the same number of observations for each individual in the longitudinal model (Hougaard 2000). In network reliability analysis, matched pairs may be used to model hot standby *vs*. primary or hot standby *vs*. cold standby. In software reliability analysis, the idea may be used to model the options such as launching a plug-in application in a separate environment or embedded within the application.

In longitudinal data model, covariates can be constant over time but individual-specific, or be time-dependent. In some studies, one can adopt covariates to make the hazard model depend on the evolution history of the process, or may introduce covariates that depend on the external factors not considered in the statistical model (Hougaard 2000). Neglected or unobserved covariates bring forward an issue of critical importance in MSA. Unobserved common covariates may be responsible for dependence between the individuals in a group. This can be captured with shared frailty; the ability to capture unobserved covariates effects is one of the most attractive features of shared frailty models.

2.3. First Hitting Times and Threshold Regression.

The first hitting times (FHT) is a mathematical concept with extensive research and wide implications in many stochastic processes, such as Wiener process, Markov chain, and counting process. The examples of FHT include lifetime, duration, and time-to-event (Lee and Whitemore, 2004, 2006). Survival data is apparently one of the FHT. In a recent survey, Lee and Whitemore (2006) reviewed this field and one of the key statistical modeling techniques for studying FHT, threshold regression. Actually, many of the familiar models in survival analysis, such as accelerated failure time, proportional hazards, competing risks analysis can be formulated as special cases of threshold regression of FHT. According to Lee and Whitemore (2006), a first hitting time model, FHT, consists of two parts: (i) A parent stochastic process $\{X(t), t \in T,\ x \in X\}$ with initial value $X(0)=x_0$, where T is *time space* and X is *state space* of the process. (ii) A boundary set $B \in X$. It is also called threshold or barrier. Assume that initially the process is outside boundary set B, the FHT is then a random variable (S), which is defined by the definition: $S = \inf\{t : X(t) \in B\}$. Both the parent process and boundary set of FHT generally depend on covariates; this is the area that the threshold regression structure is developed.

Lee and Whitemore (2006) formulated the following stochastic processes as FHT models: Bernoulli process as negative binomial first hit time, Poisson process as Erlang first hitting time, Wiener process as inverse Gaussian first hitting time, Gamma process as inverse gamma hitting time, Ornstein-Uhlenbeck process as Riccicardi-Sato first hitting time, Markov (semi-Markov) chain as first hitting time of the absorbing state. Obviously, almost all of these models are used in survival analysis. In addition, the competing risks analysis can be formulated as the latent FHT. The threshold regression is more flexible than the regression models used in survival analysis. For example, the so-called marker process, which is an external stochastic process that "accompanies" the changes of the parent process, can be incorporated into the threshold regression. The parent and marker process then form a bivariate stochastic process $[X(t), M(t)]$. Several marker processes may be combined as a single composite market process.

In reliability modeling, apparently, the shock damage model may be formulated as an FHT model. Marker processes should be very useful for modeling computer networks. For example, network reliability, performance, bandwidth, etc, can be treated as marker processes since they co-vary with network reliability. Of course, both parallel and longitudinal data models can be used in the threshold regression modeling.

3. OBSERVATION CENSORING AND TRUNCATION

Hougaard (2000) contrasted the censoring in univariate and multivariate survival analyses. The following discussion mainly draws from this reference. In univariate survival analysis, left, right, and random censorings are studied, with right censoring as the standard type. Right censoring is further distinguished as Type-I and Type-II. There is also interval censoring *vs*. point censoring. In MSA, the standard practice is to observe processes from *0* until some time C, which may be different from process to process. For the parallel data model, *homogenous* and *heterogeneous* censorings are possible. Homogenous censoring, also termed univariate censoring, may happen when the whole group is treated as a stochastic process. This is the standard for studying similar organs in biomedicine. For example, observation termination or death causes simultaneous censoring for both eyes. Given the similarity between reliability parallel systems with the studies of similar organs in biomedicine, the homogenous censoring should be important in reliability modeling. In heterogeneous censoring, each individual object may have a different censoring time.

Often censoring times and failure times are assumed independent. In some complex models, the so-called process-dependent censoring is considered. Process-dependent censoring implies that censoring at time t is dependent on the observed process up to time t. For example, in a quality assurance test for electronic bulbs, the observation is continued until a pre-specified number of bulbs are burned out. This termination time is the censoring time and is failure process dependent. Of course, in process-dependent censoring, the failure time and censored time may be positively correlated or dependent. This is the Type-II censoring (Hougaard 2000).

In recurrent events data modeling, process-dependent censoring may occur in the following manner: each

recurrence may have a negative effect on the individual, for example, each event may induce a probability p of exit from the study and hence censoring. This negative effect may be more common in reliability modeling of computer networks, where each occurrence of malicious attack increases the vulnerability of the system and thus the probability of failure. Positive effect may also be possible, this could occur in software testing processes, where the discovery of a bug may reduce the probability of the discovery of the next one. In computer networks, some events may not cause immediate failures, but lead to the performance degradation. For example, the down of an on-demand routing link only exposes the downstate when a routing request is sent. One potential use of the censoring in analyzing network faults can be to treat *omission faults* in *hybrid fault models* as process-dependent censoring. Another potential application could be the behavior modeling of the *Trojan horse* software. Once the event that triggers Trojan occurs, the software enters into a state that is of fatal damage to the system, but the consequence is not necessarily the immediate failure of nodes.

Data truncation in MSA can have more varieties than in univariate survival analysis. The simplest truncation type would be the assumption that there have been no events until the relevant time. For example, in a two components parallel system, both components are operating at the start time of observations; one ignores their previous failure or repair history. In other cases, either truncation can be due to our standards for picking individuals in the study explicitly or implicitly (Hougaard 2000). For example, an insurance company may set the life insurance premium so high for terminally ill persons to exclude them essentially. For longitudinal data, one may or may not know the histories of the individual objects, for example, the exposure to environment contamination. Another type of truncation is the unobserved or unobservable covariates; capability to considering this type of censoring is a unique feature of shared frailty modeling, one of the key motivations that inspired the development of the approach.

4. DEPENDENCE STRUCTURES AND MEASURES

Dependence is often the central focus of MSA, and together with censoring, they complicate the survival analysis most significantly. In MSA, there are several dependence mechanisms, and each may have different causes and consequences. Hougaard (2000) reviewed three mechanisms of dependence and we introduce each type based on this review: (i) common events, (ii) common risks, and (iii) event-related dependence. Common events are equivalent to common mode failures in reliability analysis, for example, failure events due to accidents or natural disasters. Usually, different objects may have different endurances to a natural disaster; otherwise, the modeling would be simple if all objects were destroyed. The common risks mechanism describes the scenario that the individuals are dependent due to some common unobserved risks, such as common genes in siblings or a bug in the operating system that may affect all the software running on it. Again, different individuals may be affected differently by the common risks because of individual differences. The third mechanism, event-related mechanism, refers to the phenomena that the actual event itself changes the risk, such as virus infection of a computer node. The risks for neighbor nodes of a virus infected computer would increase because the possibility of virus penetration via network. The first and third mechanisms are often dealt with by multi-state models and the second by frailty models in MSA. Since this article is focused on shared frailty models, which seem to lack counterparts in the engineering reliability field, the common risks dependence and frailty models are naturally the main concern in this article.

4.1. Probability Mechanisms

The term probability mechanism might be a misnomer in the sense that it does not specify the distributional assumptions that are part of the model itself. Instead, probability mechanism refers to the fashion data is generated by processes with a biological interpretation, and it actually avoids specific distributional assumptions. As explained by Hougaard (2000), probability mechanism is simply a set of *structural assumptions* that specify how a model with some features behaves, such as *independence* or *conditional independence*. The structural assumptions are often qualitative rather than quantitative to allow for the biological interpretations (Hougaard 2000). The more relevant term for the structural assumptions might be the biological mechanisms in biomedicine or design assumptions in engineering reliability.

Hougaard (2000) summarized the standard techniques of building models for multivariate random variables. One way is to formulate them as a random effects model. This corresponds to the common risks models. An alternative approach is by successive conditioning, constructing a multivariate distribution by means of the univariate distribution of T_1, the conditional distribution of T_2 given T_1 and so on. The conditional distribution approach applies to the recurrent data set, but it does not fit well with parallel data since the ordering of failure time is not specified in parallel data. Furthermore, including the time aspect introduces the event-related dependence Hougaard (2000).

4.1.1. Common Events

The term *common event* is often preserved for parallel data; for longitudinal data, the term *multiple simultaneous events* is often used. It is defined as a single cause that leads to simultaneous failures for several individuals (Hougaard 2000). In survival analysis, the most relevant are natural disasters or accidents. In software reliability, a hidden compiler bug may potentially affect all the software programs compiled with it. In semi-conductor

manufacturing, ESD (electric static damage) may damage a number of chip components simultaneously.

A practical concern for studying common events is to choose a proper time scale that corresponds to physical time for each group (Hougaard 2000). For example, to study the simultaneous deaths of married couples from accidents, a proper time scale could be the time since marriage, rather than their ages. The common events dependence does not make sense for matched-pairs data (such as those collected in a drug trial) or repeated measurements data.

4.1.2. Common Risks

Common risks model in survival analysis is the counterpart of variance components model for normally distributed data. Hence, it is a random effects model. For parallel data, the assumption is that there are some unobserved risk factors that exist across various courses from which parallel data are collected. For longitudinal data, common risks imply that the common risks factors are constant over time (Hougaard 2000). When these common risks are known, the *conditional independence* is assumed or *conditional dependence* disappears. The models are also called *latent risks models*, they are mixed models where the mixture term is common for several individuals or constants over time (for repeated events). The most common type of model for common risks is the frailty model based on a common factor in the hazards, which are expanded in later sections.

The key point with the model is that one does not know the risks deterministically, and therefore assumes that their effects are random and this creates dependence between the lifetimes. In other words, the dependence between the individuals within a group is "created" by the common risks. However, once the common risk is identified, the dependence ceases to exist, that is, the conditional independence. Consider the following example. The system administrator for a company often hesitates to upgrade the installation of their customer relations software to the latest release, because the software vendor is notorious for releasing products with bugs, especially the first release of a new version. The company CEO, however, is concerned that its major competitor may upgrade the software before his company. Therefore, the CEO leaves the decision to the system administrator but requires that no disadvantages arise from the software update decision. If both companies take the same actions, the vulnerabilities of their software installations are *dependent*, because they are exposed to the same risks from the software vendor. Since when the software is just released, one does not know the risk level and can only assume the risk is random, that is, the random effects create the dependence between two companies. However, if the system administrator has the ability to either predict or get the information by whatever reason (e.g., industrial espionage) about the status of the software release, then, conditional on the information, the software installations vulnerabilities between the two companies become independent, because the administrator can make a smart decision with the information and the information eliminates the dependence.

In general, common risks is assumed to follow a continuous distribution to reflect the notion that several risks factors exist behind the unknown risks (Hougaard 2000). In this model, the failure of one entity does not alter the risk of the others, but it influences the perception of the risks by increasing the knowledge. For example, if the system administrator's prediction is not reliable, the company may be disadvantaged. If the administrator prediction is 50/50, the information is useless.

As indicated by Hougaard (2000), it is crucial to have accounts for the known covariates. In other words, one needs to know whether a given covariate accounts part or the whole of the dependence between the individuals. The common risk may be constant over a lifetime, or be time-variant, leading to remarkably different dependence models. In some cases, an independent increments model may lead to simultaneous failures due to the introduction of time-dependence (Hougaard 2000).

4.1.3. Event-related dependence.

In this dependence model, the actual event changes the risks for future events. In computer networks, the clustered server nodes would be a typical example, since loss of one node may affect the failure probability of the other nodes in a cluster. In reliability engineering such as aerospace engineering design, the loss of redundancy is a very common issue. All event-related dependencies are not negative. For example, some diseases may induce immune responses from the patient. The chain-reaction may be a vivid description for the event-related dependence. Event-related dependence may also be used to model learning in computer science.

4.2. Dependence timeframe.

What makes dependence even more complex is that it may change with time. Furthermore, dependence changes the timeframe for individuals to be dependent (Hougaard 2000). Common events can lead to an instantaneous dependence, since failures happen at the same time. The common risks dependence typically occurs in a relatively long timeframe. In contrast, the time-dependent dependence is typically in a relatively short timeframe. In reliability analysis, the notion of time-dependent hazards is characterized by the overall bathtub curve consisting of: DFR (Decreasing Failure Rates), CFR (Constant Failure Rates), and IFR (Increasing Failure Rates). Nevertheless, there is no dependence consideration in the bathtub curve, since only one variable is involved.

In MSA, various dependence timeframes can be identified. For example, early and late dependencies refer to scenarios in which dependence happens in the early or late stages of lifetimes, respectively. Intermediate and more symmetric

dependence is also possible. A more detailed description is whether the dependence timeframe is instantaneous, short term, or long term. Instantaneous dependence means that there are multiple events happening at exactly the same time. Common events failure is often instantaneous. The multiple instantaneous failures can be treated as a single event in multi-state modeling (Hougaard 2000).

The short-term dependence is often associated with common risks dependence. Both multi-state and frailty models may be built to describe this type of dependence. In the long-term dependence interactions, the whole history is important. The long-term dependence has been formulized as a Markov extension model with limited success (Hougaard 2000). Therefore, shared frailty is particular important for studying long-term dependence.

4.3. Dependence Measure.

In univariate survival analysis, there are two types of dependences. The first type, which is often ignored, is the dependence between censoring times and survival times. The second is the dependence of failure time on the covariates, which is modeled via the proportional hazards models and its various extensions. The dependence problem in MSA is fundamentally different from that of traditional multivariate analysis. In traditional multivariate analysis, when multivariate normal distribution is assumed, the Pearson correlation (product moment correlation) only measures the linear dependence. For survival data, the marginal distributions are not normal and the dependence structure is often nonlinear. In addition, small sample experiments and semi-parametric models are the norm, rather than exceptions in survival analysis (Hougaard 2000). There are at least six dependence measures often used in MSA. Most of the measures can be defined with multivariate survivor functions and some can be estimated non-parametrically. For a detailed description of the dependence measures, one should refer to Hougaard (2000). These six dependence measures are:

(1) *Pearson Correlation Coefficients*. This is only useful for linear dependence and multi-normal distribution. It is rarely useful for MSA, but can act as a measure of deviation from the normality or linearity.

(2) *Kendall's Coefficient of Concordance* (τ). It seeks to compare the orders of survival times in the same group. For example, if one assumes that males and females have different expected lifetimes, one never directly compares male with female. One only compares individuals of the same sex. Under independence within couples, and p = 1/2, then $\tau = 0$. However, τ is invariant with both linear and nonlinear transformations.

(3) *Spearman's Correlation Coefficient*. The standard estimate of Spearman's correlation with complete data is based on the marginal ranks of survival times. This index requires continuous marginal distributions.

(4) *Median Concordance*. The idea is to avoid the conceptual shortage of Kendall's coefficient, which requires two pairs (e.g., males, females) to interpret. Instead of comparing with a second pair, one evaluates the concordance of a single observation (T_1, T_2) in relation to a fixed bivariate point. One option is to choose median lifetimes as the fixed point, and one gets the concept of median concordance. The median Concordance is:

$$\kappa = E\, sign\{(T_1 - median(T_1))(T_2 - median(T_2))\} \qquad (1)$$

or defined with bivariate survival function [$S(t_1, t_2)$] as,

$$\kappa = 4S[S_1^{-1}(1/2), S_2^{-1}(1/2)] - 1 \qquad (2)$$

Median concordance satisfies the same simple properties as the previous two coefficients. It ranges from *–1* to *1* and is *0* under independence.

It should be cautioned that for all the above coefficients, a value of zero is not sufficient to conclude independence (Houggard 2000).

(5) *Integrated Hazard Correlation*. It is similar to Spearman's coefficient and often requires numerical integration. Given $f(u, v)$ is the *p.d.f* of (u, v),

$$\rho_h = \int_0^1 \int_0^1 (\log u)(\log v) f(u,v) du dv - 1 \qquad (3)$$

(6) *Local Dependence Measures*. All above measures evaluate the global dependence and they do not address the dependence timeframes discussed earlier: the concepts of early/late, short term, long-term dependences. The following formula defines the dependence at time t, which is useful for addressing the dependence timeframes.

$$\rho(t) = S(t_1,t_2)\frac{\partial S(t_1,t_2)}{\partial t_1 \partial t_2} \Big/ \left\{\frac{\partial S(t_1,t_2)}{\partial t_1}\right\}\left\{\frac{\partial S(t_1,t_2)}{\partial t_2}\right\}. \qquad (4)$$

Another similar research area to the dependence measure is statistical tests of multivariate survival functions. Li and Lagakos (2004) investigated the properties of several important statistical tests based on the marginal analysis of multivariate survival data.

5. SHARED FRAILTY MODELS FOR PARALLEL DATA

In this section, we introduce the essential concepts, models and procedures of shared frailty models primarily based on Hougaard (2000) and supplemented with advances in the last few years. Frailty is a term used to describe the common risks, acting as a factor modifying the hazard function. The shared frailty models therefore address the common risks dependences. The approach has been developed for both parallel data and recurrent events data. This section is restricted to the parallel data. Various extensions to other data models will be discussed in sections 6 and 7.

Shared frailty can be considered a mixture model, and frailty (denoted as Y) can be treated as a mixture term. Common risks are assumed random. The model is a conditional independence model, in the sense that, given the values of the frailties, all lifetime observations become independent (Hougaard 2000). In the previous example of software updates, the dependence between the two companies is *created* by the common risks. However, once the common risk is identified, the dependence disappears. This is what is meant by *conditional independence*.

As introduced in subsection 2.1, parallel data consists of two sets of observations, one for the response times T_{ij} and the other for the individual failure indicators, D_{ij}, where i = 1, ..., n, and j = 1, ..., k. The number of individuals k may vary between groups, that is, k_1, ..., k_n, but most formulae are derived for single groups, making indexing of k unnecessary. Censoring is assumed arbitrary, either homogenous or heterogeneous.

In shared frailty modeling for parallel data, the value of Y is constant over time and common to all entities within the group. Therefore, Y is attributable to the creation of dependence, which is why the term *shared* is used. However, Y varies between groups as a random variable, which leads to different risks for the groups and the dependence between groups. A shared frailty model can be considered as a random effects model with two sources of variation: the group variation described by the random variable Y (the frailty), and the individual variation, described by the hazard function, which is denoted by $\mu(t)$ (by $\lambda(t)$ in univariate survival analysis).

As indicated by Hougaard (2000), when the frailty is strictly positive, the shared frailty model leads to absolute continuous distributions. This implies that the shared frailty model is not applicable to the common-events dependence because the common events failures create discontinuities in multivariate distributions. It is also not relevant for event-related dependence, because an event in frailty modeling may only change the information available on perceiving frailty but not the risk itself (Hougaard 2000). However, the latter point seems to be changing. Box-Steffensmeier and De Boef (2006) indicated that in the recurrent events, dependence (correlation) could result from two sources: heterogeneity across individuals and event dependence. They further suggested that the conditional frailty model of joint *event-dependence* and heterogeneity is very useful in modeling recurrent data.

5.1. Unspecified Frailty.

In the unspecified frailty model, the frailty is simply used to *separate* groups. A researcher considers to what extent Y can be freely specified—permitting each group to have a parameter specifying the risk for that group. This approach is used when the problem is to examine the effects of covariates that vary between the individuals in a group, for example, scenarios such as the *repeated measurements* study or a clinical trial with *matched pairs*. As we explained previously, in a *repeated measurement*, the time to same type of event is studied for a fixed number of times for the individuals in the experiment. In computer science, the example of repeated measurements could be the recurrences of the same type of software bugs within a specified period. The *matched pairs covariates* scheme is used in a drug trial where sets of two individuals are created on the basis of common values of some covariates (race, family), and then the members of the pair are assigned with one of the two different treatments (drug pills vs. placebos). Pairing is done by their values of key factors, in particular by the risk factors for which it is difficult to assign relevant numerical values. Similarly, the paired analysis could be designed to study the reliability of server nodes in various geographical locations. Another pairing factor could be the types of operating systems.

An intuitive idea is to treat the group factor as a regression variable, that is, define z_2, ..., z_n as indicator functions for the $n–1$ last groups. Then include them in the model, together with z_1, the indicator of the treatment group. Group *1* is excluded as a covariate, because the corresponding parameter acts as a *scale* parameter. Then a standard Cox's model (e.g., Cox 1972, Cox and Oakes 1984) can be formulated. The hazard for an individual with treatment variable z_1 is $\lambda_0(t)\exp(\beta_1 z_1)$ in group *1*, and $\lambda_0(t)\exp(\beta_1 z_1+\xi_i)$ in group i, i = 2, ..., n. However, it was found by Holt and Prentice that the standard Cox model is not appropriate; instead, the stratified Cox model by group i is more appropriate, $\lambda_i(t) = \lambda_{0i}(t)\exp(\beta_1 z_1)$, for each group, i = *1*, ..., n. There are several shortages associated with the unspecified frailty model (Hougaard 2000). In particular, the artificial designation of group *1* makes the model awkward. However, its derivation is simple and intuitive. The remainder of the section is essentially various natural extensions of frailty models to make them more realistic. The first extension is to treat ξ_i as a random variable, rather than as a parameter for each group. This is discussed in the following subsection.

5.2. General Shared Frailty Model.

In this subsection, it is assumed that there is a distribution to the frailty. The random frailty implies that one can *integrate* the frailty *out of* the expressions and thus evaluate the multivariate distribution of the response *times*. This is the main approach for parallel data frailty modeling. The major computation can be done via *Laplace transform* (Hougaard 2000).

Instead of attaching a fixed parameter ξ_i to each group (labeled as i-th group), the quantity attached is considered as a random variable in the *general shared frailty* model. To simplify the model, let us denote the quantity as $Y = exp(\xi_i)$, which will become clear later. For the individuals within a group i, the members possess independent lifetimes conditional on the values of Y_i. This is similar to the unspecified frailty case in the previous subsection, that is,

no frailty variation within a group. The key difference from the unspecified frailty model is obviously the replacement of parameter ξ_i with the random variable Y_i. The extension allows one to quantify the difference between groups readily and makes it possible to predict the responses of other individuals. Another difference is that more general regression covariates can be included to study common covariates and determine whether they are responsible for the group differences. This makes it possible to test the hypothesis that the responses times are independent when those common covariates are explained (known). The third difference is that the extension removes that unnatural treatment in the previous unspecified frailty model, measuring all risks relative to the group *1*. In the general frailty model, it is unnecessary to specify a reference group *1* any more, actually it is not permitted to do so, because the assumption of independent groups.

To preserve the validity of the independence between groups, the *scale* parameter of the distribution *Y* must be fixed. This has been a focus of significant research (Hougaard 2000). One simple solution is to eliminate the *scale* parameter, but it is only useful in very limited occasions. An alternative approach is to start by fixing the *scale* parameter in the distribution of *Y*.

Two-parameter Gamma distribution has been the dominant frailty model used in frailty modeling. However, recent research of three-parameter *generalized gamma frailty* models may change this. It is a power generalization of gamma distribution and includes other frailty distribution such as Weibull and lognormal frailty as special cases (Balakrishnan and Peng. 2006). Cox and Chu et al. (2007) presented a comprehensive tutorial on the survival analysis based on the generalized gamma distribution.

5.2.1. The mixture model and the parameterization

Mixture model— When only a single group is considered (thus group index *i* is omitted), the shared frailty model is a *common risks model* conditional on the shared frailty *Y*. The hazard function for individual *j*, conditional on the shared frailty *Y* is,

$$Y\mu_j(t) \tag{5}$$

This is a mixture model, also known as mixture distribution and it is one way to make parametric models (that are based on some standard statistical distributions, such as Exponential or Weibull) fit to data better (Hougaard 2000). A mixture model is based on a chosen parametric model, and assumes that one of its parameters varies between individuals. The parameter value is assumed to be random and follow some distribution. Furthermore, as the value is unknown, it has to be *integrated out* (in the terms of calculus). In univariate survival analysis, this approach serves two purposes. One is to generate more general distribution models, and the other is to serve as a heterogeneity model, where the population is interpreted as a mixture of individuals with different risks. The latter is also known as *compounding* (Hougaard 2000).

There are *two simple ways* to derive such mixture models for survival data. One is to impose a *scale factor* on the *hazard* (function), often denoted as λ. The other way is to impose the *scale factor* on the *survival time* (*T*), often denoted as *c*, rather than on *hazard*. One interesting fact is that with Weibull distribution, the two approaches produce the same result, but in all other cases, the results are different. The *scale factor* itself may be treated as a random variable, rather than parameter; then often *Y* is used directly, rather than using λ.

The mixture model is extremely useful for studying heterogeneous population, such as HIV carriers or some genes in a population, since separating the individuals into sub-populations may be infeasible. In reliability analysis, factors such as mixture of components from different suppliers or codes from individual programmers, etc, may be the good candidates for adopting mixture models.

To explain the two approaches, two examples are presented below. The ***first*** example is to impose *scale factor* on the hazard function. It is assumed that *Y* is a random variable representing the scaling factor. The μ(t) is determined by the standard pure hazard function. The hazard conditional on *Y* is expressed as the same equation as (5), $Y\mu_j(t)$, where $\mu(t)$ is *1* in the exponential case and $\gamma t^{\gamma-1}$ in the case of Weibull distribution. The scaling factor *Y* is the frailty in this case, describing the individual's unobserved risk.

There are quite a few confusing terminologies for *Y* in the literature, depending on the *stage* or the *information* available. Generally, the distribution of *Y* is termed the *mixture distribution*. The distribution of $\mu_j(t)$ is termed *conditional distribution*, when *Y* is fixed. The *observed distribution*, that is, when *Y* is *integrated out*, is called the *marginal distribution*. The hazard in this distribution is called the *marginal hazard* and is denoted as $\omega(t)$. A general formula for the marginal distribution can be derived by using the *Laplace transform* of the mixture distribution, that is,

$$L(s) = E[\exp(-sY)]. \tag{6}$$

The conditional survival function $S(t|Y)$ is:

$$S(t \mid Y) = \exp\{-YM(t)\} \tag{7}$$

$$M(t) = \int_0^t \mu(u)du \tag{8}$$

The marginal survival function can be evaluated by:

$$S(t) = E[S(t \mid Y)] = E[\exp\{-YM(t)\}] = L[M(t)], \tag{9}$$

where $L[M(t)]$ is the Laplace transform.

Another very important concept is the *updating of the distribution* of Y under heterogeneity interpretation of the mixture model. Since different individuals have different frailty risks, the composition or heterogeneity of a population may change over time with the progression of failure process. For example, the high-risk individuals exit the system faster than low-risk individuals do. One of the relationships between the marginal and condition hazard is,

$$\omega(t) = \mu(t)E[Y \mid T > t], \tag{10}$$

which indicates that the observed hazard (marginal) is equal to the mean hazard, when evaluated among the survivors at the time of study.

The ***second*** way to build a mixture model for survival data is to *scale* the failure or response times. This is also known as *time-scale mixture model*. The following is an example. Let $c = Y^{-1/\gamma}$. The lifetime T can be described as cW, where W is Weibull $(1, \gamma)$ distributed. This relation can be immediately used to evaluate moments and other properties, for example, $E[T] = E[c].E[W]=E[Y^{1/\gamma}]E[W]$.

A more *general time-scaled mixture model* can be defined simply as: $T = ZW$, where Z and W are positive valued random variables, with Z acting as a random variable for the *scale factor c*.

Conditional parameterization—refers to the relaxation that allows the conditional hazard [$\mu_j(t)$] to take on arbitrary or non-parametric expressions. That is, in the $Y\mu_j(t)$, $\mu_j(t)$ is allowed to be non-parametric, and the values of Y is common to individuals within a group.

The approach needs to be extended to *multivariate survival distributions*. Independence of the *lifetimes* implies no variability in the distribution of Y, that is, when Y has a *degenerate distribution*. When the distribution is not degenerate, the dependence is positive. In a few cases, the model can be extended to allow for negative dependence. For simplicity, let us begin with the bivariate case to explain the extension.

Conditional on shared frailty Y, the bivariate survival function is:

$$S(t_1, t_2 \mid Y) = \exp[-Y\{M_1(t_1) + M_2(t_2)\}], \tag{11}$$

where $M_j(t) = \int_0^t \mu_j(t)du$, $j = 1,2$ are the integrated conditional hazards. From (11), one derives the bivariate survival function by *integrating out Y*.

$$S(t_1, t_2) = E\exp[-Y\{M_1(t_1) + M_2(t_2)\}] = L[M_1(t_1) + M_2(t_2)] \tag{12}$$

$E(.)$ is the expectation and $L[.]$ is the *Laplace transform* of the distribution of Y. The bivariate survival function is then expressed via the *Laplace transform* of the frailty distribution, evaluated at the total integrated conditional hazard. Similarly, let

$$M(t_1,...,t_k) = \sum_{j=1}^{k} M_j(t_j),$$

the multivariate survival function for k observations is:

$$S(t_1,...,t_k) = L[M(t_1,...,t_k)] \tag{13}$$

From the survival function, the *p.d.f.* can be derived by differentiation with respect to $t_1,...,t_k$, and the likelihood function can then be built. As to the hazard function $\mu_j(t)$, various choices can be made, and the simplest choice is the *symmetry* $\mu_j(t) = \mu(t)$, all hazards are equal. Another option is that they are unrelated, or arbitrary, independent of each other. There is also a compromise between the all equal and arbitrarily independent, that is, the proportional hazards, $\mu_j(t) = \rho_j\mu_0(t)$, where the ρ_j is a set of parameters for j=1, ..., k observations. Finally, the Cox's proportional hazards form can be used, with, $\mu_j(t) = \exp(\beta z_j)\mu_0(t)$.

Furthermore, various frailty distributions from exponential, Gamma, Weibull, PVF (power variance function) family may be utilized for different applications.

Geoffrey and Rocke (2002) found that when double censorings exist, the commonly used *partially likelihood estimation* cannot be used. They suggested using the conditional independence in the model as much as possible. The survival function for an individual with frailty Y_i is defined as: $S_j(t \mid Y_i) = \exp[-Y_i(\rho_j t)^{k_j}]$ for the control group and $S_j(t \mid Y_i) = \exp[-\tau Y_i(\rho_j t)^{k_j}]$ for the treatment group. Therefore, the treatment is assumed to have a multiplicative effect on the conditional hazards, $h_j(t|Y_i)$. (Here, we changed Geoffrey and Rocke's (2002) notation for frailty from Z_i to Y_i to keep consistent with the rest of the article).

Marginal parameterization— Marginal parameterization is an alternative to the conditional parameterization. That is, all formulae are expressed with the marginal distributions. The advantage for the alternative is that there is an approximate orthogonality between parameters, which makes it easier to estimate the quantities (Hougaard 2000). The derivation can be conducted via the marginal survival functions $S_j(t)=\Pr(T_j>t)$, defined by $S_1(t)=S(t, 0)$, $S_2(t)=(0,t)$, ..., etc, or the model can be based on the marginal hazard function, that is the hazard in the marginal distribution, $\omega_j(t)$, with integrated hazard $\Omega_j(t)$. The marginal survivor function and hazard are related in the traditional definition fashion, $S_j(t) = \exp\{-\Omega_j(t)\}$. From the relation, $S_j(t) = L[M_j(t)]$, it can be derived that, $M_j(t) = L^{-1}[S_j(t)]$. The bivariate survival function, corresponding to Equation (12) is:

$$S(t_1, t_2) = L[L^{-1}(S_1(t_1)) + L^{-1}(S_2(t_2))] \tag{14}$$

The multivariate version of survival function is drawn similarly as, corresponding to (13),

$$S(t_1,...,t_k) = L[\sum_j L^{-1}(\exp\{-\Omega_j(t_j)\})] \qquad (15)$$

The Equation (15) can be differentiated to obtain multivariate *p.d.f.*, which can be used to build likelihood function for parameter estimations.

5.2.2. Frailty Distribution Updating

Similar to the discussion in the *mixture model section*, the updating for frailty distribution is necessary. The underlying reason for updating is that after a period of observations, either with or without some events occurred, one gets more information about the frailty *Y*. This is due to the heterogeneity phenomenon, which implies that high-risk group experienced some events, while low-risk groups have experienced fewer or no events. The updating is essentially the re-evaluation procedure for the conditional distribution of *Y,* given the observed quantities since the start of the experiment (Hougaard 2000).

*Updatin*g can be used to assess the effects of truncation since truncation may influence the distribution of events in a population with heterogeneous frailty. *Updating* can also be harnessed to evaluate conditional distributions, if the objective is to make individual forecasting. The so-called *dynamic evaluation*, where the hazard at each time instant is calculated based on all the information known up to that time, or the filtration (F_{t-}), is obviously extremely meaningful. Under the dynamic approach, the hazard of death for the *j-th* individual given the individual is alive at time *t* is:

$$\lambda_j(t \mid F_t) = \mu_j(t)E(Y \mid F_{t-}) \qquad (16)$$

This is actually a generalization of the Equation (10).

Truncation may alter the measure of dependence (e.g., Kendall's τ). The updating model can be used to quantify the change (Hougaard 2000). Updating mechanism can be used to simulate the effects of *malicious truncation* on survivability of computer networks.

5.2.3. Quantification of dependence.

According to Hougaard (2000), generally, three ways of expressing dependence can be used in a frailty model. Firstly, some measure of frailty variability should be adopted, given that the dependence itself is possibly time-variant due to the frailty heterogeneity. Often the variance of Log(*Y*) is used, if the objective is to evaluate dependence via some measure of variability of the frailty distribution. The second approach is to evaluate dependence by a correlation type measures as defined in section 4. The third approach is to evaluate the conditional distributions. In the traditional multivariate analysis, this corresponds to analyzing the linear regression relation. In the *multi-state* models, a dynamic version of the relation is used. In frailty models, it is handled by the dynamic updating formulae as discussed in previous paragraph. The third approach offers the result as a whole curve, rather than a single number, due to the including of dynamic updating scheme (Hougaard 2000).

5.2.4. Regression models

In frailty modeling, it is required to consider explanatory covariates because the frailty describes the effects of common unknown factors. If some common covariates are incorporated in the model, the variation owing to unknown covariates should drop (Hougaard 2000).

Assume there are *p* observed covariates z_{ijm}, $i = 1,..., n$, $j=1,...,k$, $m=1,...,p$, for each individual. Often, they are denoted as *p* vectors z_{ij}. Generally, they depend on *i* and *j* in an arbitrary way. Two special cases are the common covariates and the matched-pairs covariates. Common covariates are common for all members of the group, that is, depends on *i* only. Matched pairs covariates are covariates that only depend on *j*.

The traditional proportional hazards model can be used to incorporate covariates, but need to express as conditional on the distribution of frailty *Y*. In fact, the frailty can be interpreted as a common but unobserved risk factor, which produce the following conditional hazard function for the (*i,j*)-*th* individual,

$$\mu_j(t)\exp(\beta' z_{ij} + \psi'\omega_i) \qquad (17)$$

where ω_i is the vector of common unobserved covariates and the hazard term $\mu_j(t)$ may or may not depend on *j*. Let $Y_i = \exp(\psi'\omega_i)$, a conditional hazard of

$$\exp(\beta' z_{ij})Y_i\mu_j(t) \qquad (18)$$

is obtained as the standard formulation.

5.3. Shared Frailty with Specific Frailty Distribution

In previous section, the general frailty model is introduced. We only stated that frailty follows some distribution, but did not give specific distribution. In this section, we mention a few distributions that have been used frequently.

Gamma Frailty model: — The two parameters Gamma (δ, θ) distribution, with θ as *scale* parameter has been utilized in frailty modeling to generate mixtures distributions from the very early days of frailty modeling. It possesses quite a few advantages, such as the simplicity of the derivatives of the *Laplace transform*, easy computation of the Kendall's τ [$\tau = 1/(1+2\delta)$]. In addition, updating with Gamma distribution is easy, since the distribution of *Y* among the survivors at time *t* is also Gamma-distributed with the

parameter δ unchanged and θ changed to $\theta + \sum_j M_j(t)$. It is the original distribution modified by a scale parameter. This is a unique property of Gamma distribution. Under Gamma frailty distribution model, if the conditional hazard follows Weibull distribution, the bivariate survival function takes the form of *Burr distribution* that is a generalization of the well-known *Pareto power distribution*.

The gamma frailty model can be extended to describe negative dependence, which may be rare but does occur in practice. The negative dependence implies that the death of one individual may actually lower the risk of the other individual. One example in computer networking would be the death of a node that is being utilized as the base for attacking other nodes or a node that has become the source of virus spreading.

PVF (Power Variance Function) frailty distributions — The PVF family of distributions is a natural exponential family and has the property that variance and mean satisfy power-law model. The previous mentioned Gamma, positive stable distributions, and the inverse Gaussian distributions are the special cases of PVF family. The PVF has three parameters, α, δ, θ. In the case of $0<\alpha\leq 1$, it can be obtained as the distribution of Y in a stable frailty model after truncation. The special case $\alpha=\delta$, $\theta=0$ leads to positive stable model. The case of $\alpha = 0$ derives the gamma distribution. When $\alpha = 1/2$, the mixture distribution is an inverse Gaussian distribution. For $\alpha < 0$, there is a *point mass* at zero, implying that some groups correspond to zero risk. It is a desirable feature for capturing the individuals of immunity or that some groups being unable to experience the event considered (Hougaard 2000). For example, in computer network modeling, one may use the feature to capture the cross-platform immunity—a virus written for Windows® may not affect a Linux node.

Other frailty distributions — Theoretically, any distribution of positive random variables may be utilized as frailty distribution models. One key point is the simplicity of the *Laplace transform*. Others distributions that have been studied for frailty modeling include: *positive stable frailty*, *lognormal*, *inverse Gaussian*, *generalized inverse Gaussian*, *Frank's* distributions. In particular, recent studies suggest that the *positive stable frailty distribution* should be preferred to the Gamma model in scenarios when the correlated survival data exhibit a decreasing association with time (Martinussen and Pipper 2004). Martinussen and Pipper (2004) developed a likelihood estimation procedure for the positive stable shared frailty Cox model. Ravishanker and Dey (2000) used the finite mixtures of positive stable frailty distributions to form a multivariate survival model. They used the *cross-ratio function* as a local measure of dependence.

5.4. Statistical inference for shared frailty models.

Despite the omission of the statistical inference of the frailty models here, its importance is self-evident. Without feasible estimation procedures, the models simply cannot be applied to the practical data. One basic notion is to *integrate out* the random frailties, but there are other alternatives. As reviewed by Hougaard (2000), one alternative is to interpret the frailty as unobserved random variables, similar to the best linear unbiased predictor (BLUP) method for normal distribution models, and then EM-algorithm can be used. The challenge with EM-algorithm, which is reasonably simple, is the possibly excessive number of iterations required. The EM-algorithm includes both the frailty (Y) and the observed quantities (T, D) into estimation procedure, giving a full likelihood of $L_{(T,D)|Y}L_Y$, where the first term is the standard survival likelihood given the frailties and the second term is the likelihood from frailty density (Hougaard 2000).

The so-called three-stage approach described by Houggard (2000) is another major methodology. This method is advanced to address the problems that a parametric model may not give satisfactory fit to the marginal distribution, the alternative semi-parametric approach, which combines parametric models for dependence with classical non-parametric estimates for the marginal distribution. Markov chain Monte Carlo (MCMC) has also been used for the gamma frailty model, which simulate the distributions of frailty values, avoiding the difficulty in handling the complex likelihood functions. Besides the parameter estimation procedures, goodness-of-fit testing methods are also necessary. Asymptotic theory for these models is still not fully developed. Commercial software packages such as SAS®, SPSS®, have not yet released programs for all frailty modeling. Vu and Knuiman (2002) proposed an approach based on maximum likelihood and EM to address the parameter estimations in the semi-parametric marginal shared gamma frailty models. Gorfine et al. (2006) developed a so-called *pseudo full likelihood* approach that is able to handle general frailty distribution with finite moments and achieved similar efficiency as the widely used EM algorithm. Ripatti et al. (2002) devised a maximum likelihood inference procedure for multivariate frailty models based on an automated Monte Carlo EM algorithm.

5.5. Spatial Frailty Modeling.

Back to 2001, Li and Ryan (2001) indicated that, "*there has been, however, virtually no literature dealing with models for spatially correlated survival data*". This seems still largely true. They proposed was a new class of semi-parametric frailty models extended to process the spatially correlated survival data. Spatially correlated data exist widely in practice, despite the lack of statistical methods specifically tailored for them. For example, epidemiological data in public health or the animal population dynamics data in fields are full of spatial dependence or correlations.

The following is an extremely brief sketch of Li and Ryan's (2001) model. Assume there are M geographical regions,

and each region has n_i objects being observed. For each of the objects, there are: $X_{ij} = \min(T_{ij}, C_{ij})$, where T_{ij} are failure times and C_{ij} are censored times. There is also a corresponding covariates Z_{ij} for each individual object. In addition, each region is assigned a region-specific random effect $r(p_i)$, where p_i represents the specific region. When survival time T_{ij} is assumed independent, the conditional hazards function (termed intensity function in original paper) is of the following form:

$$\lambda_{ij}\{t \mid Z_{ij}, r(p_i)\} = \lambda_0(t)\exp[\beta' Z_{ij} + r(p_i)] \quad (19)$$

where β is the fixed effects vector and $r(.)$ assumes the form of stationary Gaussian process with zero mean.

In a very recent study, Bastos and Gamerman (2006) proposed what they called *dynamic model with spatial variation*, and the model is not limited by the proportional hazard assumption. The model is expressed with the hazard function $h(t, X, s)$,

$$h(t, X, s) = \exp[X'\beta(t) + Z + W(s)] \quad (20)$$

where t is the time, s is the space, X is the vector of covariates, $\beta(t)$ is the time-dependent vector of regression coefficients. $W(s)$ is the spatial frailty for space location s. Z is the unstructured spatial frailty. Several models such as Cox (1972) proportional hazard model, the original frailty model by Clayton (1978), Henderson et al's spatial frailty (2002), Carlin and Banerjee's (2002) dynamic survival, can be derived from the model as special cases.

6. SHARED FRAILTY FOR OTHER DATA MODELS

Shared frailty models for the parallel data, discussed in previous section 5, is the most extensively studied field in frailty modeling, and it also has the longest history. However, frailty modeling approach has been extended significantly to other survival data in recent years. In particular, it has been extended to the data sets of *recurrent events*, *short-term and instantaneous frailty*, and true *multivariate frailty models* (beyond bivariate). Given the extensive contents and also the similarities between these other extensions with the frailty models for parallel data, we only mention the most fundamental differences between the extensions. In particular, we omit all the equations and models, which seem necessary to fit an article of this length but make it nearly impossible to have comprehensive discussion on the relevant topics here. Readers are referred to Hougaard's (2000) excellent monograph for the detailed treatments. The following is mainly summarized from several chapters of Hougaard (2000).

6.1. Shared frailty models for recurrent events

The recurrent events data is significantly different from parallel data. In the parallel data model, one pre-specifies a group of individuals or items, which are followed individually until failure. In recurrent events model, one is interested in single individuals who experience the same event multiple times. Recurrent data is essentially longitudinal data. Recurrent data may be modeled as multi-state model by treating the number of occurrences of recurrent events as different states. The discovery of bugs in a software-testing project is an example of recurrent events data in computer science.

In the case of independent data, the *event count* can be described with the ordinary Poisson process. This means that future events are independent of previous ones, which requires that there are no random variations among the individuals. In the more general cases with dependence, one treats the models conditional on some individual's variations (frailties). Thus, the frailty model for recurrent events is a model for dependence or individual variation (Hougaard 2000).

The frailty models for recurrent events are similar to those for parallel data discussed in section 5. Here are the major conceptual differences summarized by Hougaard (2000): The frailty variation in recurrent events modeling does not exist as group variation, but exists as variation between individuals. This is opposite to the interpretation in parallel data in which frailty is considered as a random variable taking different values between groups. Furthermore, the variation described by the hazard function is not an individual's variation, but a variation between individuals, which is called the Poisson variation in recurrent events modeling. For parallel data set, the risk decreases at each event time (since high risks individuals dies early); whereas for recurrent events data, the risk set is constant over the observation period. Therefore, it is critical to observe the times of events for parallel data, whereas for recurrent events, the frailty leads to the variation in the number of events, even the observation time is the same for all individuals. In the extreme case of very long observation period, every individual in parallel data observation will ultimately die, so there is no variation in the number of events (death). However, obviously the variation is inevitable in the case of recurrent events. Therefore, in recurrent events data, it is satisfactory to know the number of events within the observation period and the ultimate actual times is less concerned. In software testing, the discovery of bugs is a typical recurrent event. One depends on the number of bugs discovered within the testing period for decision-making, but depending on the ultimate actual number of bugs may simply be impractical.

When the observation period is the same for all individuals, some evaluations can be based on only the number of events during the study period (the acquired data is called *period count data*.). This leads to specific *Poisson over-dispersion* distributions and a completely separate theory to study them.

The stochastic process is often Markov for frailty models in recurrent events, whereas the process for multi-state bivariate data is often not Markov. The reason for the difference is that the risk set in recurrent events is constant over time. Accordingly, the constant hazards assumption often holds for recurrent event data, but is rarely valid for parallel data. For parallel data, dependence measures such as Kendall's coefficient are sufficient, but for recurrent event data, we need fundamentally different measures. Poisson Process, over-dispersion models such as Negative Binomial distribution, PVF distributions, and regression models based on Cox model extensions are the major models for describing recurrent frailty.

Peña (2006) reviewed a set of new dynamic process models based on counting stochastic process and the Martingale central limit theorem in the context of recurrent events. This class of models, which is able to handle the unique key features (such as mentioned in section 2.1) of recurrent events simultaneously, has been developed by Pena and his collaborators in the last few years (Peña and Hollander 2004, Stocker and Peña 2007). The potential application fields of this class model envisioned by the authors include survival analysis, reliability and maintenance. The model emphasizes the dynamic nature of reliability and shows the necessity of stochastic process such as counting process and Martingale theory in recurrent event modeling.

6.2. Short-term and instantaneous frailty models.

Dependence can be distinguished into three timeframes: long-term, short-term, and instantaneous. Almost all the dependence models so far deal with long-term dependence. Multi-state models can be easily converted into short-term dependence modeling by abandoning the Markov assumption and substituting with a Markov extension model. However, the resulting models often lack interpretation, and are hard to solve analytically. In contrast, frailty model is relatively ameanble when extending to the short-term dependence modeling (Hougaard 2000). Similar to the previous subsection 6.1, the following is simply a list of the topics; readers are referred to Hougaard (2000) monograph for more details.

Frailty models are essentially random effects models with three sources of variations. In shared frailty models for parallel data (Section 5), these are group effects and individual random effects, plus random group by time interaction effects. In recurrent events, the corresponding random effects are individual variation and simple variation. In the context of short-term and instantaneous dependence, one supplements the recurrent modeling treatment with a random variation over time. Both parallel and recurrent events are involved in the short-term and instantaneous timeframe modeling. According to Hougaard (2000), the following seven types of models have been developed to model short-term and instantaneous dependence.

(i) *Independent increments frailty model,* which deals with simultaneous events and instantaneous dependence. This describes common risks models in a randomly changing environment. Compared to the shared frailty models, it substitutes the constant frailty Y, with a stochastic process common to all individuals in the group $Y(t)$.

(ii) *Bivariate parallel data* model. It is assumed that the individuals in a group share the same realization of $Y(t)$, and the groups are assumed independent and have the same distribution of the stochastic process.

(iii) *Recurrent events data model*. Whether instantaneous dependence can occur for recurrent events data is case dependent, since in some cases, it is physically impossible to experience two events simultaneously. Even when the physically simultaneous events are impossible, the model may still be useful as an approximation to the scenarios that dependence is extremely short-term. Generally, the instantaneous dependence model is simpler than that for short-term dependence, but the latter is often closer to practice. The *increments frailty* modeling can be used to build models for the recurrent events data, the model is actually simpler than that for parallel data since the risk set is constant. An alternative to increment frailty modeling is called *subordinated time model*, which assumes the *increments of frailty* as a stochastic *time process* in continuous time starting at time 0. For example, the *time process* may be a Poisson process that can describe the accumulated use of the object. In the case of computer networks, the time process can describe the repeated attacks from hackers or the repeated calls of a module in software.

(iv) *Piecewise gamma model.* This model is formulated in a similar way as the *additive model*. What is derived from the additive frailty over time structure is then a model with *piecewise constant frailty*.

(v) The so-called *moving average model* is intended to "smoothen" the additive frailty model, that is, to reduce the jumps in the dependence in the piecewise constant frailty model.

(vi) The *hidden cause of death model* is still a shared frailty model and is based on the idea that each person has a constant, but multivariate frailty, with one coordinate for each cause of death. Similar to the moving average model, this also makes the change smoother over time.

(vii) *Woodbury-Manton diffusion process model* is a model that may possess extremely rich features. It was developed in the context of univariate survival analysis. Due to the computational difficulties even in the univariate case where numerical solution is required, its applicability is unknown in the multivariate case (Hougaard 2000). The model is a stochastic differential equation and it treats the frailty as a diffusion process controlled by a Brownian motion.

As pointed out by (Hougaard 2000), a major issue with the above models is their mathematical complexity. The extreme difficulty is that even simplified versions may still be too complex and cost computationally to be applicable.

7. MULTIVARIATE FRAILTY MODELS

In this section, similar to section 5, we present a brief introduction to the key issues in multivariate frailty models, largely based on Hougaard (2000), and supplemented with some of the latest advances in literature. The shared frailty models in previous sections are somewhat limited in the sense that the dependence is treated in a pair-wise fashion. This is perfectly appropriate for bivariate data with common risks dependence. For general multivariate systems with three or more failure variables, the ideal models should be able to consider *varying degree* of dependence. The theory for generalization of shared frailty modeling to general multivariate systems is still an actively research field. Only *ad hoc* extensions are available. As indicated by Hougaard (2000), the main mathematical issue lies in that some very essential probability results on mixtures modeling in multivariate context are not yet developed.

There are two general requirements for a multivariate frailty model. The *first* is the *conditional independence*, that is, conditional on the frailty. The frailty may be a random variable (univariate or multivariate) or a stochastic process. It is also assumed that the object being studied behave independently; this implies that the frailty capture all the dependence between response times. The previous frailty model is essentially a common risk model, the multivariate frailty model is more general and it also includes the features such as negative dependence, which gives flexibility to capture both negative and positive interactions between objects. This is a progress with significant potential in practical applications. Put a broad perspective, it provides a general framework for modeling interactions and dependence well beyond the applications of survival analysis or reliability, such as machine learning, cooperative systems modeling, since negative dependence in failures implies positive interactions in survival or cooperation. We see that the prospect for expanding to other subjects is extremely promising. For example, Locatelli et al (2007) formulated a bivariate frailty model to study woman's susceptibility to breast cancer by considering genes, environment, as well as immunity under the unified frailty model, the *correlated frailty-mixture* model. The immunity was incorporated with negative dependence mechanism, as expected. We believe that the approach can also be applied to fields as diverse as animal population demography and dynamics, software reliability, and network performance modeling. The *second* requirement for multivariate frailty model is that the frailties act multiplicatively on the hazards, like what was adopted in previous sections, such as $Y\lambda(t)$. This restriction is not mandated technically (Hougaard 2000).

The data considered here is parallel data taking the form $T_1,..., T_k$ in a single group. The overall modeling strategy is to replace the shared frailty Y with multivariate random variables (Y_1, ..., Y_k), such that Y_j applied to time number j. The previous shared frailty model is then the special case when $Y_1 = ... = Y_k$ and the distribution of Y is the same for all groups. In this section, the frailty is still assume *constant*; the time-varying frailties are too complex for multivariate frailty modeling at this stage and seem not explored yet (Hougaard 2000).

In parametric modeling of multivariate survival analysis, one has to specify the multivariate probability distribution on $(0, \infty)^p$. This is often performed by generalizing the univariate distribution to multivariate counterparts, but sometimes the generalization could be troublesome due to the extra complexity in high dimensions. Walker et al (1999) introduced a univariate family of distribution, termed the *beta-log-normal family*, which can be extended to multivariate system naturally. This new family is motivated by the mixture of some of the typical distributions.

7.1. Differential effects of a shared frailty.

One extension of the shared frailty model is to allow the various response times being influenced by different frailty functions. That is, the hazard for individual j is in the form of $f_j(Y)\mu_j(t)$ in time t. Define $Y_j = f_j(Y)$, such that the hazard is expressed as: $Y_j\mu_j(t)$. When $f_j(Y), j = 1,...k$ is the same for all j and does not depend on j, the model defaults to the standard shared frailty model. Often the f_j are monotone, and one can define the model via the first frailty term $Y_1=Y$, as some kind of base frailty. In proportional hazards modeling, a natural choice is the power functions, because they fit into $Y_j = \exp(\psi_j'\omega)$, where ω represents the originally neglected covariates which are treated as constant, but the corresponding coefficients ψ_j are allowed to vary and therefore the Y_j is different for each j. This can be interpreted as that frailty has different degree of influence on different responses, or it is more important for some responses than for others. In the medical context where this model was first developed, this formulation was proposed simply as a mathematical interpretation and the practical implication was not agreed upon (Hougaard 2000). If we think this interpretation in the engineering context, it may indeed possess important practical implications. Treat j as different events, for example, in a computer network, the events can be malicious compromise, hardware failure, natural degradation such as OS memory leaks, etc. This model will allow us to assign different frailty parameters Ψ_j for different events, all the events occur under the same covariates (ω) which could represent maintenance or environment, such as on the same OS patches, or same firewall protection. However, the effects of the frailty on various events apparently can be very different. For example, software maintenance has strong effects on

malicious compromises but little effects on hardware failures.

7.2. The multiplicative stable model.

If X_1 and X_2 are independent and X_1 follows positive stable distribution with parameter α and X_2 follows the same distribution with parameter ρ, then the distribution of $X_1^{1/\rho}X_2$ is positive stable with (αρ). This derives a multiplicative mixture model, with the distribution staying within the family of positive stable distributions. The results allow one to evaluate some *multivariate Laplace transforms* for models with nested data.

The following is an example of the so-called *nested trivariate parallel data model* of T_1, T_2, and T_3, which may be used for modeling siblings. Assume T_2 and T_3 are twins, and T_1 is an ordinary sibling. Naturally, T_2 and T_3 are strongly dependent, but T_1 is less dependent with the twins. There are three frailties Y_1, Y_2, Y_3 generated from powers of three independent variables: Z_0 is common to all individuals, Z_1 is for individual *1*, Z_2 is applicable for individuals *2* and *3*. That is,

$$Y_1 = Z_0 Z_1, \qquad Y_2 = Y_3 = Z_0 Z_2 \tag{21}$$

It is assumed that Z_0^{ρ} follows the positive stable distribution with parameter α and is denoted as *PosStab*(α), and Z_1 and Z_2 are denoted as *PosStab*(ρ). To fit the mixture model described in the beginning, Z_0 corresponds to $X_1^{1/\rho}$. It follows that all the marginal distributions are identical. The marginal distribution of each Y is *PosStab*(αρ) . This formulation allows that both the conditional and marginal distributions of lifetimes are common to the three individuals. However, the degree of dependence is higher between T_2 and T_3 than between T_1 and T_2 or T_1 and T_3.

What was derived is a random effects model with three sources of variation. Two of these sources, the group effect (Z_0) and the individual variation captured by the hazard $\mu(t)$, are similar to the corresponding quantities in the shared frailty model. The third (represented by Z_1 and Z_2) model a source common to only a subgroup (twins) of the group under study.

The conditional hazards model for the *j-th* individual, can be represented as

$$\lambda_0(t)\exp(\beta' z + \psi' w + \phi' u_j) \tag{22}$$

The covariates effects are expressed as three parts: z is the observed part, w is the common unobserved covariates, and u is the unobserved covariates common to individuals *2* and *3* only. That is, $Z_0 = \exp(\psi' w)$ and $Z_j = \exp(\phi' u_j)$, j=1,2. The survivor function for the trivariate distribution is then:

$$S(t_1,t_2,t_3) = \exp(-[M_1(t_1)^{\rho} + \{M_2(t_2) + M_3(t_3)\}^{\rho}]^{\alpha}) \tag{23}$$

If the conditional distributions follow Weibull distribution, this model can be formulated alternatively via the accelerated failure times model.

7.3. Additive models

The above multiplicative model is unique to the *positive stable distributions*. In general, it is much more convenient to handle additive frailty models (Hougaard 2000). Here, the bivariate frailty is used as an example to demonstrate the additive models. Assume that the conditional hazard for individual j is of the form $Y_j \mu_j(t)$, given the bivariate frailty (Y_1, Y_2). The model is constructed by using the following additive functions of random variables:

$$Y_1 = Z_0 + Z_1 \text{ , } Y_2 = Z_0 + Z_2, \tag{24}$$

where Z_0, Z_1, Z_2 are independently distributed variables. Of course, Z_0 is introduced to capture the common frailty between Y_1 and Y_2, since it contributes to both Y_1 and Y_2. Z_1 and Z_2 are independent individual terms, generating only extra variance.

The model can be interpreted as a *hidden cause of death model*. Suppose that there are two causes of death, with proportional hazards, $c_1\mu(t)$ and $c_2\mu(t)$, and corresponding frailties Y_1 and Y_2. Further assume cause No. *1* is genetically determined, but cause No. *2* is unrelated to genetics. It could be, that cause Y_1 is shared by family members, whereas cause Y_2 is individual, that is, it is in the form of Y_{2j} for individual j in a family. When the actual death is unknown, this gives the additive model, with Z_0=c_1Y_1, Z_1=c_2Y_{21}, Z_2=c_2Y_{22}. The proportional hazards are crucial for formulating the relative simple additive model here, but extending to non-proportional hazards is also possible (Houggard 2000).

The conditional survivor function in the general bivariate distribution case is,

$$S(t_1,t_2 \mid Z_0,Z_1,Z_2) = \exp\{-(Z_0+Z_1)M_1(t) - (Z_0+Z_2)M_2(t_2)\}, \tag{25}$$

where M_1(t) and M_2(t) cumulative or integrated hazards function. The bivariate survivor function can be expressed via *Laplace transforms*, $L_m(s)$ for Z_m, m=0,1,2 as follows:

$$S(t_1,t_2) = L_0[M_1(t_1) + M_2(t_2)]L_1[M_1(t_1)]L_2[M_2(t_2)] \tag{26}$$

The *p.d.f.* can be derived by differentiation with respect to t_1 and t_2, from which likelihood functions can be built.

The gamma distribution for Z has been exclusively adopted in the *additive frailty models* to make the marginal distributions of Y_j simple. For example, by imposing the common value for the parameter θ of gamma distributions

on all Z_m, m=0, 1, 2, the resulting marginal distribution is also gamma. The model is also called *correlated frailty model* in literature (Hougaard 2000).

Quite a few additive models based on the above framework have been proposed. We mention one interesting model, the *father-mother-child* model. The father-mother-child model is inspired by the genetic relationships between parents and child, where some genetic terms are common for the father and child and some for the mother the child. There are also environmental terms common to all of them in the family. One particular point about the model is the so-called common marginals. For example, if there is a frailty term for father and child, we need an independent contribution for the mother with the same distribution. Five items Z_1 to Z_5 are required to guarantee that the model leads to common marginals. The model can be described by a matrix model corresponding to the linear equations system, the frailty model is:

$$\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix} = \begin{pmatrix} Z_1+Z_2+Z_5 \\ Z_1+Z_3+Z_4 \\ Z_1+Z_2+Z_4 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} Z_1 & Z_2 & Z_3 & Z_4 & Z_5 \end{pmatrix}^T \tag{27}$$

This matrix equation specifies the model for frailties, analysis similar to the above bivariate case can be conducted but the process is much more complex (Houggard 2000). Intuitively, this model should provide a useful framework for studying the reliability of software built with the objected-oriented approach.

As summarized by Hougaard (2000), current multivariate shared frailty modeling tries to incorporate more complicated dependence structure over individuals, but not more complication over time. The multiplicative models possess many nice theoretical characteristics. However, the additive models seem relatively easier to apply. The dependent structures are very flexible in additive models. The key computational issue in applying additive models is to obtain the multivariate *Laplace transforms*. A disadvantage with the additive model is that one needs more parameters to make models, which requires data sets that contain more detailed information (Hougaard 2000).

7.4. Proportional Odds Models

In biomedical research, the proportional odds models have been proposed to be an alternative to the proportional hazards model. Let T_{ij} denotes the failure time of the j-th observation from the i-th group (j=1, ...n_i and i=1, 2,...n). n is the number of groups (also known as clusters) and n_i is the size of the group i. Corresponding to each T_{ij} is the covariates z_{ij} and an unobserved latent variable U_{ij} which induces the intra-group dependence. The groups are independent. Conditional on the random effects u_{ij} and u_{ih}, T_{ij} and T_{ih} ($j \neq h$) are mutually independent. The multivariate proportional odds model can be formulated in several ways, one of which is the conditional hazard function of T_{ij}, and given $h'(t)$ is the first derivative of $h(t)$,

$$\lambda_{ij}(t \mid u_{ij}, z_{ij}) = \frac{h'(t)\exp\{h(t) + u_{ij} + z_{ij}^T\beta\}}{1+\exp\{h(t) + u_{ij} + z_{ij}^T\beta\}} \tag{28}$$

The ratio of any two conditional hazard functions converges to unity over time, $\lambda_{ij}(t)/\lambda_{ih} \to 1$, is a key property of the proportional odds models. This implies that the heterogeneities between several groups diminish over time. For example, the difference between patients under some treatments and individuals in controls should diminish overtime. Ultimately, the patients recover totally due to the treatment and demonstrate no difference from healthy controls. The proportional odds model can be extended to multivariate survival data, Lam and Lee (2006) provided such an example. Lu and Zhang (2007) discussed the covariate selection in the proportional odd model.

We argue that the notion of unit hazards ratio can be used to compare reliability of two dynamic systems or to compare the reliability of repaired systems with old systems. System maintenance and medical treatments are analogically the same; therefore, the modeling approach should be useful in engineering reliability.

7.5. Proportional Mean Residual Life Models

This model was first proposed by Oakes and Dasu (1990), Chen and Cheng et al. (2005) derived the semi-parametric inference procedure by using the counting process theory. The mean residual lifetime is,

$$m(t) = E(T - t \mid T > t) \tag{29}$$

The proportional mean residual life model by Oakes and Dasu (1990) is:

$$m(t \mid Z) = m_0(t)\exp[\beta^T z] \tag{30}$$

$m(t|Z)$ is the mean residual lifetime, conditional on the covariates , and $m_0(t)$ is baseline mean residual lifetime.

7.6. Other Issues in Multivariate Shared Frailty Modeling

Multivariate lognormal frailty model — The simplest way to build a multivariate frailty model is to use a multivariate lognormal distribution. For example, in the bivariate case, the frailty (Y_1, Y_2) is given by (X_1, X_2) = (log Y_1, log Y_2), then (X_1, X_2) follows bivariate normal distribution. However, it is difficult to do explicit computation, as simple expression for the *Laplace* transform is not available. Numerical integration, penalized likelihood, etc can be used to deal with the computation. There have been few applications of this model. Jensen (1998) applied a Bayesian lognormal frailty model to study the genetic relations between bulls and conducted computation with Gibbs sampling.

Negative dependence models—Positive dependence is dominant in survival analysis. However, negative dependence does exist and may be of crucial importance in some cases. For example, negative dependence is very likely to occur in the recurrent events data from an *alternating state process*. *Shared resource competition* is another scenario where negative dependence occurs when the strong competitors died first. First, there must be procedures to detect whether the dependence is positive, negative or neutral. This can be accomplished by extending an existing model with positive dependence to accommodate for negative dependence. Second, we need models for the cases of negative dependence. The shared gamma frailty model can be extended to accommodate negative dependence, but its probability density could be too concentrated, especially when the dimension increases. Multivariate normal frailty can also be extended to negative dependence. Aalen (1987) proposed a distribution where (Y_1, Y_2) is the diagonal in a *Wishart* distributed matrix. In most cases, the model gives a positive dependence, but for some parameter values, the dependence is negative.

MFPT approach —This is a recently proposed approach to relax the limitation of traditional Cox's proportion hazards assumption by Sauerbrei et al. (2007). They used the MFPT (*multivariable fractional polynomial time-transformation)* to select variables with influences on the outcome, determine a sensible dose-response function for continuous variables, and investigate time-varying effects on a continuous scale. The motivation is to extend the MFP (*multivariable fractional polynomial)* approach by Sauerbrei and Royston (1999) to accommodate the nonlinear covariates effects in the Cox proportional hazards approach.

Parsimonious analysis of time-dependent effects in the Cox model—Lehr and Schemper (2007) warned about the over-fit of Cox proportional hazards model and suggested to adopt the parsimonious strategy in data analysis in Cox framework. They noted that as long as sufficient degree of freedom is spent, the fitting of any shape of time-dependence is possible, but the resulting model may be an over-fitted curve. The over-fit will inevitably lead to the increased width of confidence intervals and decreased prognostic power. They demonstrated that *fractional polynomials* and similarly *penalized likelihood* approaches are suitable in controlling the over-fit by parsimonious use of degrees of freedom, still allowing flexibly exploring the time-dependence contained in data sets.

Time Process Regression—Fine et al. (2004) recently proposed a general linear model for survival time *vs*. covariates that are allowed to be temporal process observed over intervals. This model includes extension of standard models in multi-state survival analysis but it does not require smoothing or a Markov assumption. Therefore, it is different from the currently widely used multi-state models that are based on the transitional probability (intensities in survival analysis). The following is an extremely brief sketch of the model formulation: Let $Y(t)$ be the mean of lifetime at time t, $Y(t)$ is conditional on a vector of time-dependent covariates vector $Z(t)$ and a time-dependent stratification factor $S(t)$. Then, the model is defined as:

$$E(Y(t) \mid z(t), S(t) = 1\} = g^{-1}\{\beta^T(t) z(t)\} \tag{31}$$

where g is a monotone, differentiable and invertible function. For example, g^{-1} may take the form of logistic function such as $g^{-1} = \exp(.)/[1+\exp(.)]$. $\beta(t)$ is the vector of time-dependent coefficients. Obviously, in the previous section, the regression coefficients vector is not time-dependent. This is a fundamentally different feature from the other models. These parameters are interpreted conditionally on covariates at t, not all $s \le t$. In addition, the Y(t) is the mean of response (survival time) and therefore does not involve a Markov assumption. There is also an indicator function $\delta(t)$ [R(t) was used in original paper, which we change to avoid confusion with reliability], set to *1* when the temporal process is observed, and to *0* otherwise. The function g(.) is also called *link function*, a term from general linear modeling and g^{-1}(.) is termed inverse link function. In a recent paper, Rabe-Hesketh and Skrondal (2007) applied the general linear modeling to derive a framework for multilevel and latent variable modeling, with the approach of *composite links* and *exploded likelihoods*. The derived models are envisioned to be applicable in survival analysis.

Bootstrap analysis of multivariate failure data — It is interesting to notice that survival analysis appears to have been dominated by semi-parametric approaches, and pure nonparametric approaches have been used much less. Monaco et al (2005) demonstrated that Bootstrap method could be used to estimate standard errors in the multivariate failure times, in particular when the concern is the point failure probability.

8. SUMMARY AND PERSPECTIVES

Shared frailty modeling is suitable to common-risks *dependence type*, parallel, repeated measurements, and recurrent events *data models*, as well as long-term, short-term, and instantaneous *dependence timeframes*. In addition, recent research seems to suggest that event-related dependence can also be described with frailty models. Obviously, shared frailty theory occupies the central focus of multivariate survival analysis. The dominance is not accidental because frailty captures crucial but often directly unobservable (latent) random variables, which are attributable for dependence and variations. To some extent, *dependence*, *variation*, *randomness*, and *frailty* all address some facets of multivariate failure events. The frailty theory provides a unique set of approaches to address the theoretical and practical issues related to those four concepts. Similar to the fact that censoring is the trademark feature of survival analysis (both univariate and multivariate), frailty theory is the trademark of multivariate

survival analysis. These two trademark properties, censoring-handling and frailty modeling, have made multivariate survival analysis a unique and indispensable mathematical tool for biomedical and public health research.

The applicability of survival analysis and its sister field competing risks analysis to engineering reliability is out of question. Obviously, there have been early interactions in these fields. What is perhaps unfortunate is that survival analysis and reliability diverged significantly in recent years and their interactions withered accordingly. In our opinion, this happened in a period when survival analysis has already achieved significant breakthroughs. This had apparently been fueled by the demands from biomedical research, which has been undoubtedly the most active subject in the last quarter of the 20th century. In contrast, during the same period, industrial engineering reliability was perhaps perceived to be mature and sufficient despite the existence of some crucial issues such as dependence.

Another interesting observation is that there is significant theoretical research in both survival analysis and reliability fields, mostly by mathematicians. Some of these theoretical studies have been published in reliability literature and some are in survival analysis, but the majority has been published in mathematical and statistical journals. Although different terms are used, the problems addressed are similar, and the solutions are often very similar too. What seems to make the difference is that in survival analysis fields, the high demands from biomedical research helped to convert the theoretical research into statistical procedures and methods. Similar demands from reliability engineering to reliability theory seem not as strong as in biomedicine.

It is our opinion that due to this background today's survival analysis can offer some unique opportunities to the engineering reliability field. Frailty modeling is the most significant in our opinion, because it addresses the most difficult issue in both fields—dependence. Therefore, for reliability and survival analysis, our motivation is to bridge the diverged gap. In addition, we also believe that the interaction should be bidirectional. One convincing example is multi-state modeling, where the reliability field should be able to offer fresh insights to the survival analysis field.

In computer science, and to a lesser extent, in the IEEE related engineering fields, multivariate survival analysis is still relatively unknown. As we have argued in the first two articles of the series, survival analysis and competing risks analysis should be able to address many crucial issues in *network reliability* and *survivability*, *software reliability* and *test measurements* (Ma and Krings 2008a,b). This is certainly also the case for frailty modeling, which we believe should provide more powerful modeling techniques, as we have commented in various occasions in the previous sections. For example, the negative dependence in the frailty models is not that important for survival analysis. However, the modeling of negative dependence may open significant applications in computer science such as machine learning and cooperative systems modeling.

In the previous two articles, we also argued the potential of survival analysis and competing risks analysis in prognostic and health management (PHM), software reliability, and network survivability. We believe frailty modeling provides an even more compelling case because of its multivariate nature and its unique advantage in handling dependence.

REFERENCES

Aalen, O. O. 1987. Mixing distribution on a Markov chain. *Scand. J. Statist*. 14:281-289.

Aalen, O. O. 1994. Effects of frailty in survival analysis. *Statistical Methods in Medical Research.3:227-243.*

Andersen, P. K., O. Borgan, R. D. Gill, N. Keiding. 1993. Statistical Models based on Counting Process. Springer.

Balakrishnan, N. and Y. Peng. 2006. Generalized gamma frailty model. *Statist. Med*. 2006; 25:2797–2816

Bastos, L. S. and D. Gamerman. 2006. Dynamic survival models with spatial frailty. *Lifetime Data Analysis*. 12:441-460.

Box-Steffensmeier, J. M. and S. De Boef. 2006. Repeated events survival models: The conditional frailty model. *Statist. Med*. 2006, 25:3518–3533

Carlin, B. P, Banerjee S (2002) Hierarchical multivariate CAR models for spatio-temporally correlated survival data. In: Bernardo J. M. et al, eds. Bayesian Statistics. Oxford Univ. Press, pp 1–15

Chen, Y. Q., S. Cheng. 2005. Semi-parametric regression analysis of mean residual life with censored survival data. Biometrika (2005), 92, 1, pp. 19–29

Chiang, C. L. 1991. Competing risks in mortality analysis. *Annual Review of Public Health,* 12 281–307.

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

Commenges, D. 1999. Multi-state models in Epidemiology. *Lifetime Data Analysis*. 5:315-327.

Cox, D. R. 1972. Regression models and life tables. *J. R. Stat. Soc. Ser. B*. 34:184-220.

Cox, D. R. & D. Oakes. 1984. *Analysis of Survival Data*. Chapman & Hall. London.

Cox, C., H. Chu, et al. 2007. Parametric survival analysis and taxonomy of hazard functions for the generalized gamma distribution. *Statist. Med.* (in Press)

Crowder, M. J. 2001. Classical Competing Risks. Chapman & Hall. 200pp

David, H. A. & M. L. Moeschberger. 1978. The theory of competing risks. Macmillan Publishing, 103pp.

Escobar, L. A. and W. Q. Meeker. 2006. A Review of Accelerated Test Models. Statistical Science. 2006, Vol. 21, No. 4, 552–577.

Fine, J. P., J. Yan., M. R. Kosorok 2004. Temporal process regression. Biometrika 91(3):683–703

Fleming, T. R. & D. P. Harrington. 1991. Counting process and survival analysis. John Wiley & Sons. 429pp.

Geoffrey, J., and D. M. Rocke. 2002. Multivariate survival analysis with doubly-censored data: application to the assessment of Accutane treatment. Statist. Med. 2002; 21:2547–2562

Gorfine, M., D. M. Zucker. 2006. Prospective survival analysis with a general semiparametric shared frailty model: A pseudo full likelihood approach. *Biometrika* (2006), 93, 3, pp. 735–741.

Harrell, F. E., Lee, K. L. et. al. 1996. Multivariate prognostic models: issues in developing model, evaluating assumptions and adequacy, and measuring and reducing errors. *Statist. Med.* 15,361-387 (1996)

Henderson R, Shimakura S, Grost D (2002) Modelling spatial variation in Leukaemia survival data. *J Am Stat Assoc* 97:965–972

Hougaard, P. 1999a. Multi-state Models: A Review. *Lifetime Data Analysis*. 5:239-264.

Hougaard, P. 1999b. Fundamentals of Survival Data. *Biometrics*. 55:13-22.

Hougaard, P. 2000. Analysis of Multivariate Survival Data. Springer. 560pp.

Ibrahim, J. G., M. H. Chen and D. Sinha. 2005. Bayesian Survival Analysis. Springer. 481pp.

Jafar, S., A. W. Krings and T. Gautier. 2008. Flexible Rollback Recovery in Dynamic Heterogeneous Grid Computing, IEEE Trans. on Dependable and Secure Computing (*in print*).

Kalbfleisch, J. D. & R. L. Prentice, 2002. The Statistical Analysis of Failure Time Data. Wiley-InterScience, 2nd ed. 462pp.

Klein, J. P. and M. L. Moeschberger. 2003. Survival analysis techniques for censored and truncated data. Springer.

Kvam, P. and Peña, E. (2005). Estimating load-sharing properties in a dynamic reliability system. *J. Amer. Statist. Assoc.* 262–272.

Lam, K. F., and Y. W. Lee. 2006. Merits of Modeling Multivariate Survival Data Using Random Effects Proportional Odds Model. *Biometrical Journa*l 46 (2004) 1, 331–342

Lawless, J. F. 2003. Statistical models and methods for lifetime data. John Wiley & Sons. 2nd ed. 630pp.

Lee, M. T. and G. A. Whitmore. 2006. Threshold Regression for Survival Analysis: Modeling Event Times by a Stochastic Process Reaching a Boundary. Statistical Science, 2006, Vol. 21, No. 4, 501–513

Lee, M. T. and Whitmore, G. A. (2004). First hitting time models for lifetime data. In Advances in Survival Analysis (C. R. Rao and N. Balakrishnan, eds.) 537–543. North-Holland,

Lehr, S. and Michael Schemper, M. 2006. Parsimonious analysis of time-dependent effects in the Cox model. *Statist. Med*. 2007; 26:2686–2698.

Li, Y., and L. Ryan. 2001. Modeling spatial survival data using semi-parametric frailty models. *Biometrics*. 58:287-292

Li, Q. 2004. Comparison of test statistics arising from marginal analyses of multivariate survival data. *Lifetime Data Analysis,* 10:389-405.

Lisnianski, A., Levitin, G. (2003). Multi-State System Reliability: Assessment, Optimization and Applications. World Scientific Press.

Liu L., R. Wolfe, and X. Huang. 2004. Share frailty models for recurrent events and a terminal event. *Biometrics*. 60:747-756.

Lin, D. Y. 1994. Cox regression analysis of multivariate failure time data, the marginal approach. *Statist. Med*. 13:2233-2247.

Locatelli, I., A. Rosina, P. Lichtenstein and A. I. Yashin. 2007. A correlated frailty model with long-term survivors for estimating the heritability of breast cancer. *Statist*. *Med*. (in press)

Lu, W, and H. H. Zhang. 2007. Variable selection for proportional odds model. *Statist. Med*. 2007; 26:3771–3781

Ma, Z. S. 1997. Survival analysis and demography of Russian wheat aphid populations. Ph.D. dissertation, 307pp, University of Idaho, Moscow, Idaho, USA.

Ma, Z. S. and A. W. Krings. 2008a. Survival Analysis Approach to Reliability, Survivability and Prognostics and Health Management. Proceedings of 2008 IEEE-AIAA AeroSpace Conference. BigSky, Montana, March 1-8, 2008. (*In Press, in the same volume*)

Ma, Z. S. and A. W. Krings. 2008b. The Competing Risks Analysis Approach to Reliability Survivability, and Prognostics and Health Management. The 2008 IEEE-AIAA AeroSpace Conference. BigSky, Montana, March 1-8, 2008. (*In Press, in the same volume*)

Ma, Z. S. and A. W. Krings. 2008c. Multivariate Survival Analysis (II): Multi-State Models in Biomedicine and Engineering Reliability. 2008 IEEE International Conference on Biomedical Engineering and Informatics (BMEI 2008). May 27th-30th, 2008. (*Accepted*)

Martinussen, T. and C. B. Pipper. 2004. Estimation in the positive stable shared frailty Cox proportional hazards model. *Lifetime Data Analysis*. 11:99-105.

Monaco, J, J. Cai, and J. Grizzle. 2005. Bootstrap analysis of multivariate failure time data. *Statist. Med*. 2005; 24:3387–3400.

Oakes, D. & Dasu, T. 1990. A note on residual life. *Biometrika* 77, 409–10.

Oakes, D. 2001. *Biometrika* Century: Survival Analysis. *Biometrika* (2001) 88(1):99-142.

Peña, E. and Hollander, M. (2004). Models for recurrent events in reliability and survival analysis. In Mathematical Reliability: An Expository Perspective (R. Soyer, T. Mazzuchi et al., eds.) 105–123. Kluwer, Boston.

Peña E. A. 2006. Dynamic Modeling and Statistical Analysis of Event Times. *Statistical Science*. 21(4):487–500

Rabe-Hesketh, S. and A. Skrondal. 2007. Multilevel and laten variable modeling with composite links and exploded likelihoods. *Physchometrika*. 72(2):123-140.

Ravishanker, N. and D. Dey. 2000. Multivariate Survival Models with a Mixture of Positive Stable Frailties. Methodology and Computing in Applied Probability, 2(3): 293-308, 2000.

Ripatti, S., K. Larsen., J. Palmgren. 2002. Maximum Likelihood Inference for Multivariate Frailty Models Using an Automated Monte Carlo EM Algorithm. *Lifetime Data Analysis*, 8:349–360.

Sauerbrei, W. and Royston, P. (1999). Building multivariable prognostic and diagnostic models: Transformation of the predictors by using fractional polynomials. *Journal of the Royal Statistical Society, Series A: Statistics in Society* 162, 71–94.

Sauerbrei, W. P. Royston, and M. Look. 2007. A New Proposal for Multivariable Modeling of Time-Varying Effects in Survival Data Based on Fractional Polynomial Time-Transformation. *Biometrical Journal* 49 (2007) 3, 453–473

Stocker, R. and Peña, E. (2007). A general class of parametric models for recurrent event data. *Technometrics*. In Press.

Vu, H. T. V. and M. Knuiman. 2002. Estimation in semi-parametric marginal gamma frailty models. *Aust. N. Z. J. Stat*. 44:489-501.

Walker S. G. and D. A. Stephens. 1999. A multivariate family of distributions on $(0, \infty)^p$. *Biometrika*. **86(**3):703-709

Wei, L. J. and D. V. Glidden. 1997. An overview of statistical methods for multiple failure time data in clinical trials. *Statist. Med*. 16:833-839

BIOGRAPHY

Zhanshan (Sam) Ma holds a Ph.D. in Entomology and is a Ph.D. candidate in Computer Science at the University of Idaho. He has published approximately 30 journal and 30 conference papers, mainly in the former field. Prior to his recent return to academia, he worked as senior network/software engineers in software industry. His current research interests include reliability and survivability of wireless sensor networks, fault tolerance, survival analysis, evolutionary game theory, evolutionary computation and bioinformatics.

Axel W. Krings is a professor of Computer Science at the University of Idaho. He received his Ph.D. (1993) and M.S. (1991) degrees in Computer Science from the University of Nebraska - Lincoln, and his M.S. (1982) in Electrical Engineering from the FH-Aachen, Germany. Dr. Krings has published extensively in the area of Computer & Network Survivability, Security, Fault-Tolerance and Real-time Scheduling. In 2004/2005 he was a visiting professor at the Institut d'Informatique et Mathématiques Appliquées de Grenoble, at the Institut National Polytechnique de Grenoble, France. His work has been funded by DoE/INL, DoT/NIATT, DoD/OST and NIST.