Modeling longitudinal data of congestive heart failure patients: a case study at Wachemo University Nigist Ellen Mohammed memorial Hospital

A Thesis Submitted to the Department of Statistics, College of Natural Science, Jimma University as a Partial Fulfillment for the Requirements of Masters of Science (MSc.) Degree in Biostatistics

November; 2018
Jimma, Ethiopia

Modeling longitudinal data of congestive heart failure patients: a case study at
Wachemo University Nigist Ellen Mohammed memorial Hospital

MSc thesis

By: Mohammed Sultan

Advisor: Mr. **Dechasa Bedada (MSc.   Ass.proff.)**

Co-advisor: Mr. **Birhan Worku (MSc.)**

October; 2018

Jimma, Ethiopia

Jimma University

College of Natural Science

Department of Statistics

Modeling longitudinal data of congestive heart failure patients: a case study at Wachemo University Nigist Ellen Mohammed memorial Hospital

By: Mohammed Sultan

As members of the board of examiners of MSc thesis open defense examination of the above title we read and evaluate the thesis and examined the candidate.

Advisor   Signature   Date

Mr. **Dechasa   Bedada (MSc., Ass.prof)**

Co-advisor   Signature   Date

Mr. **Birhan Worku (MSc.)**

Internal examiner   Signature   Date

External examiner   Signature   Date

Chairman   Signature   Date

ACKNOWLEDGMENT

ABSTRACT

Background of the study: Congestive heart failure (CHF) is a complex clinical syndrome that can result from any functional or structural cardiac disorder that impairs the ventricle's ability to fill with or eject blood. There are different vital signs of CHF from those most commonly Heart rate, respiration rate, and weight monitoring in the follow-up to assess the progression of congestive heart failure disease. These markers are correlated and needed to ensure an accurate evaluation of them since each has its own limitations and could be influenced by demographical and physiological characteristics of the patient.

The objective of the study: The main objective of this study was modeling longitudinal data of congestive heart failure patients in a Case study at Wachemo University Nigist Ellen Mohammed memorial Hospital.

Methods: In this study secondary data was used from Wachemo University Nigist Ellen Mohammed memorial Hospital in CHF Outpatient Clinic. The study consists of 154 CHF patients, measured repeatedly at minimum three and maximum nineteen times on each patient who is 18 years old or older for those visited Hospital from December 2015 to January 2018. The linear mixed model was applied in this study to model the three outcomes of CHF.

Results: The baseline mean and standard deviation of Pr, Rr, and Wh are 106.16 and 25.37, 31.53 and 11.44 and 64.68 and 10.12 respectively. From the different correlation structure for the separate, bivariate, and multivariate model; modeling with autoregressive order one correlation structure is appropriate for CHF data in addition to unstructured covariance structure for random effects to consider within and between patients variations.

Conclusion: Finally a multivariate model was considered as best to study the joint evolution and identify the potential risk factors affecting the three end-points.

Key Words: Modeling; Longitudinal Data; Correlation structure; Linear Mixed Model;

Table of Contents

List of tables

List of figures

List of acronyms

ACF	Acute Coronary failure
AIC	Akaike Information Criterion
AU	Africa Union
BIC	Bayesian Information Criterion
CAD	Coronary artery disease
CHD	Coronary heart disease
CHF	Congestive heart failure
GEE	Generalized estimating equations
GLM	Generalized Linear Model
Hr	Heart rate
HF	Heart failure
ICC	Intra-class Correlation Coefficient
LMM	Linear mixed model
LVEF	Left-ventricular ejection fraction
MLIRT	Multilevel item response theory
MLMM	Multivariate linear mixed model
MQL	Marginal Quasi-Likelihood
NYHAC	New York heart association class
PQL	Penalized Quasi-Likelihood
Pr	Pulse rate
Rr	Respiration rate
SBP	Systolic blood pressure
DBP	Diastolic blood pressure
Sqrtim	square root of time
Wh	Weight

1. Introduction

1.1. Background

Congestive heart failure (CHF) is a complex clinical syndrome that can result from any functional or structural cardiac disorder that impairs the ventricle's ability to fill with or eject blood [1]. This decrease in cardiac output, the amount of blood that the heart pumps, is not adequate to circulate the blood returning to the heart from the body and lungs, causing fluid (mainly water) to leak from capillary blood vessels. This leads to increased pulmonary venous pressure and fluid accumulation in CHF [2] for this reason, I) the lungs become stiffer, II) Respiration becomes insufficient, leading to hypoxemia and acidosis, III) the effort required to breathe and frequency of breaths increases. Among the vital signs, the following are three commonly measured and monitored to know the status of patients:

Respiration rate: Abnormal respiratory rates and changes in respiratory rate are broad indicators of major physiological instability, and in many cases, Rr is one of the earliest indicators of this instability. Therefore, it is critical to monitor Rr as an indicator of patient status. Rr performs at least as accurate in identifying patients at risk of these adverse events as PR and the SBP. A Rr of greater than 24 breaths per minute is able to identify approximately 50% of patients at risk of serious adverse events with 95% specificity. Although the main function of the respiratory system is a gas exchange, a broad range of factors can affect ventilation. In patients with CHF, an increase in Rr can warn of impending pulmonary edema, or fluid in the lungs, which is a common debilitating symptom of CHF as it was stated in American Heart Association [3].

Heart rate: It is among the many vital signs (respiration rate, blood oxygen saturation, arterial blood pressure, etc.), one of the most commonly measured and monitored. Whatever will be the sensing principle or the monitoring method used, data referred to the HR can be considered the primary vital sign information which is needed on a patient approach in both emergency and clinical situations. *Gorgas* [4] stated that; Hr data are used to measure anomalous rate or irregular Pr (arrhythmias) or heart block. The Hr or Pr represents the number of times the heart beats in a certain period of time. It is usually measured in minutes, and normal resting Hr is approximately 60 to 80 beats per minute. It can go as high as 100 in a healthy adult and as low as 40 in athletes as it is described in *American Heart Association and Gorgas* [3, 4]. A lower Hr is

associated with a stronger and healthier heart. A lower Hr means the heart is not pumping or working hard to deliver blood and oxygen to the body. The heart speed increased to compensate for its failing ability to adequately pump blood throughout the body. Patients may feel a fluttering in the heart (palpitations) or a heartbeat that seems irregular or out of rhythm. The pulse can be lowered through regular exercise, and there are also breathing exercises to lower the heart. Take slow deep breaths to lower the pulse.

Weight: Steady gain in weight shows that the body is retaining fluid. As blood flow out of the heart slows, blood returning to the heart through the veins backs up, causing fluid to build up in the tissues. The kidneys are less able to dispose of sodium and water, also causing fluid retention in the tissues [5]. Because the heart is not pumping blood efficiently, fluid can build up, causing swelling in the ankles, feet, legs, and belly. This fluid buildup can also cause a weight gain of more than 2 pounds in a day or 5 pounds in a week [6]. Patients are allowed to monitor their own weight each time to adjust their intake of medication, water, or salt according to their weight change. Consequently, patients' weight becomes stable within the relatively normal range and thus improves CHF prognosis.

Monitoring vital signs are a necessity for patients who underwent congestive heart failure to assess the progression of congestive heart failure disease. Most commonly the above three markers are monitored and measured repeatedly over time in the follow-up to ensure that good quality of life and long life expectancy for the congestive heart failure patients. These markers are correlated and needed to ensure an accurate evaluation of them since each has its own limitations and could be influenced by demographical and physiological characteristics of the patient [7]. Given the interdependence of heart rate, respiratory rate and weight of CHF patients in determining congestive heart failure, it is important to evaluate the factors that affect the rate of change in these outcomes in a joint manner. The motivation of this study was to assess the association between the three outcomes (Rr, Hr, and Wh) and key demographic and clinical factors accounting for the correlation between these markers in a multivariate fashion.

Analysis of multiple outcomes and longitudinal data are given special attention in the recent literature. In a different study analyzing the multiple outcomes jointly considered as the best model. For example, multivariate linear mixed models (MLMM) were proposed to analyze multiple outcomes to assess and test for global exposure effect across outcomes while assuming

a flexible correlation structure for the multiple outcomes [8]. The idea was to provide robust estimates for the mean by separating it from the correlation parameters. Multivariate linear mixed models (MLMM) can be used to account for repeated measures in longitudinal studies and also for random effects. It describes the relationship between two or more continuous response variables and independent variables, with fixed effects and random effects. Unlike multivariate general linear mixed model, it has assumptions of normality. This assumption is important that, deviation from the normality assumption could affect the accuracy of the estimates and the inferences for the high-level outcomes [9]. Multivariate joint modeling was also an alternative wherein a joint distribution is specified to jointly model all random effect [10, 11]. For congestive heart failure patients the main symptoms which explained in the above were three continuous variables. The model is a multivariate mixed effect model, this model contains fixed effects, random effects, and repeated effects. When we use a mixed effect model our interest was not limited only on fixed effects but have flexible variance-covariance structure and correlation structure.

1.2. Statement of the Problem

Congestive heart failure is one of the chronic diseases, which is a growing public health problem in both developed and developing countries. It is characterized by a variety of unpleasant outcomes, such as poor Quality of life, recurrent hospitalization, high mortality and significant cost burden (12). According to *Cook C, 2014* (13) the World Bank estimates the global economic cost at \$108 billion per annum. In 2015 heart failure affected about 40 million people globally [14]. Overall around 2% of adults have heart failure [15] and in those over the age of 65, this increases to 6–10% [16][17] Above 75 years old rates are greater than 10% [15].

Congestive heart failure is a serious long-term condition that will usually continue to get slowly worse over time. So have to know the evolutions of congestive heart failure that how to change over time and whether the patterns of the rate of changes are different or not with different covariates. Have to know also how the associations between markers evolve over time or how their evolution associate over time.

Different kinds of literature were done in the area of congestive heart failure disease. Some of the literatures were done by using a linear mixed model including random effects and independent measurement error for the outcome variables as an example *Fissuh and Muleta* [7].

The first gap in the previews study was using only two symptoms (Pr and Rr) of congestive heart failure to relate associated risk factors. The study of congestive heart failure is based on its symptoms; mean that managing symptoms is managing congestive heart failure. To manage congestive heart failure assessing different symptoms are needed as much as possible. In this study, one more symptom of CHF was added which is weight. Analyzing the model with three variables (Pr, Rr, and Wh) jointly fit best.

The second gap was independent measurement error assumption; mean that the serial correlation is accounted by variance-covariance of random effects. It may be reasonable when the within individual measurements are far apart so that within-individual autocorrelation is practically negligible, or that the between-individual variation is dominant. The independent measurement error is the default for the linear mixed model; if within-subject correlation structure is not specified.

The correlation between the measurements within the subject usually depends on the time interval between the measurements and decreases as the length of the interval increases. When the repeated measures are collected close in time or correlations among the repeated measures doesn't change at a short time, random effects alone may not adequately account for the dependency due to the repeated measures. Ignoring the existing correlation of longitudinal data may lead to incorrect and inefficient inferences, and it may increase the risk of Type I error rate and underestimating standard errors. For congestive heart failure patients restricting within-subject independent covariance structure was not appropriate. Because the follow-up time is a short and congestive heart failure is a chronic disease so measurements may not change at a short time. By considering those issues the multivariate model was fitted.

The following major research questions were addressed:
1. How the Pr, Rr, and Wh of CHF patients evolve over time?
2. What factors predict the evolution of Pr, Rr, and Wh of CHF patients separately?
3. What factors predict the joint evolution of Pr, Rr, and Wh of CHF patients?
4. Which model best fit the association between the evolution of Pr, Rr, and Wh of CHF patients and risk factors?

1.3. Objectives of the Study

1.3.1. General Objectives

The main objective of this study was to model longitudinal data on congestive heart failure patients of Wachemo University Nigist Ellen Mohammed memorial Hospital

1.3.2. Specific Objectives

The specific objectives are to:

➢ Explore the evolution of Pr, Rr, and Wh of CHF patients over time separately.

➢ Fit a separate mixed effect model for the Pr, Rr, and Wh of CHF patients and identify the associated factors.

➢ Fit joint mixed effect models for Pr, Rr, and Wh of CHF patients and identify the associated factors

➢ Compare separate, joint bivariate and multivariate models

1.4. Significance of the Study

The results of this study would be useful in the development of an effective care and patient monitoring system on chronic and/or non-communicable disease. Specifically:

➢ It will initiate the researchers to joint model multivariate longitudinal data.

➢ It will help how to identify the performance of the different joint models of multivariate longitudinal data.

➢ It will be used as a reference for those who want to apply separate and joint modeling techniques in two longitudinal continuous sequences.

➢ It will help to identify the potential risk factors influencing the separate as well as the joint evolution of respiratory rate and heart rate measurements in congestive heart failure patients. This will, in turn, help the respective policy makers of the health sector in the effort to design an appropriate intervention strategy.

2. Literature review

2.1 Empirical review

2.1.1 Description of Congestive heart failure

Heart failure, sometimes known as congestive heart failure, occurs when your heart muscle doesn't pump blood as well as it should [18]. It is a complex clinical syndrome that can result from any functional or structural cardiac disorder that impairs the ventricle's ability to fill with or eject blood [19].

Rheumatic heart disease is the most common cause of cardiac disease in general and Congestive Heart Failure in most sub-Saharan African countries, followed by hypertensive heart disease which is rising along with other non-communicable diseases. However, the pattern of congestive heart failure in our setting is not known [20].

People with Heart Failure May Experience: Breathlessness during activity (most commonly), at rest, or while sleeping, which may come on suddenly and wake you up; coughing that produces white or pink blood-tinged mucus; swelling in the feet, ankles, legs or abdomen or weight gain; a tired feeling all the time and difficulty with everyday activities, such as shopping, climbing stairs, carrying groceries or walking; a feeling of being full or sick to your stomach; memory loss and feelings of disorientation; heart palpitations; which feel like your heart is racing or throbbing[21].

2.1.2. Factor Associated with the Congestive heart failure

There are two mechanisms of reduced cardiac output and heart failure: systolic dysfunction and diastolic dysfunction. The most common causes of systolic dysfunction (defined by a left-ventricular ejection fraction of greater than or equal to 50%) are ischemic heart disease, idiopathic dilated cardiomyopathy, hypertension, and valvular heart disease. Diastolic dysfunction (defined as a dysfunction of left-ventricular filling with preserved systolic function) may occur in up to 40–50% of patients with heart failure, it is more prevalent in women, and it increases in frequency with each decade of life. Diastolic dysfunction can occur in many of the same conditions that lead to systolic dysfunction. Based on the descriptive statistics of [19], the most common causes are hypertension, ischemic heart disease, hypertrophic cardiomyopathy, and restrictive cardiomyopathy. *American Heart Association*, [22] stated as the most common cause of heart failure is coronary artery disease (CAD). CAD occurs when arteries that supply

blood to the heart muscle become narrowed by buildups of fatty deposits called plaque. Their descriptive statistics also stated as the other common risk factors that lead to heart failure are: Past heart attack has done some damage to the heart muscle, Heart defects present since birth, High blood pressure, Heart valve disease, Diseases of the heart muscle, Infection of the heart and/or heart valves, Abnormal heart rhythm (arrhythmias), Being overweight, Diabetes, Thyroid problems, Alcohol or drug abuse, Certain types of chemotherapy.

2.2. Methodological review

2.2.1. Longitudinal Data Analysis

Longitudinal data are a series of measurements of the same event taken from the same individual repeatedly over time. The most unique characteristic of longitudinal data is the ability to directly study change. The primary goal of most longitudinal studies is to characterize the change in response over time and the factors that influence this change [23].

2.2.2. Theory of the Linear Mixed Model

Analyses of multiple observations measured on the same individual over time are different from observations measured on different people. Investigators gather repeated measures or longitudinal data in order to study the change in a response variable over time as well as to relate these changes in explanatory variables over time [24].

The linear mixed-effects model fits the mean response as a combination of population characteristics (fixed-effects) assumed to be shared by all individuals and subject-specific effects (random-effects) that are unique to a particular individual [25]. By including random-effects in the model, linear mixed-effects models are able to explicitly distinguish between within-subject and between-subject sources of variation. With a linear mixed-effects model it is not only possible to estimate parameters that describe how the mean responses change over time, but it is also possible to predict how an individual's response trajectories change over time.

In longitudinal data analysis, when subjects are followed over time, there is a natural ordering of the data for each subject. Correlation structure is used to model dependence among observations, in the mixed-effect model; it is used to model dependency among the within-group errors [26].

Most researchers when using an LMM tend to assume the level one residual structure follows an independence structure without taking into account the type of data (i.e. cross sectional or longitudinal data) as an example *Fissuh and Muleta* [7]. He fitted the bivariate linear mixed model with including random effects and independent measurement error. This may be chosen due to the parsimonious nature of the independence model or the researcher believes that including more random effects adequately accounts for the dependency due to repeated measures. However, the following question must be asked, after removing the variation due to the random effects are the within individual residuals independent from one another within an individual [27]? In other words, conditional on the random effects, is it tenable to assume that the within individual residuals are independent? This assumption may not hold in some data situations, especially if the time between observations is very short (i.e. daily or weekly observations) or if the correlation between observations does not decrease very quickly [27].

Lloyd J. Edwards [28] stated as an extremely important fact regarding measurements repeated on an individual is that the measurements are typically correlated. Though it could happen that repeated measurements on an individual may not be correlated, it is unlikely that repeated measurements on the same individual will actually be independent. If the correlation is ignored, it can negatively impact parameter estimation, hypothesis testing, and efficiency of study design. If the correlation is ignored, the computed confidence interval could be much smaller than the nominal level, hypothesis tests can have a much higher Type I error, and statistical power can be lower than planned. The correlation matrix and/or covariance matrix between observations play an important role in the analysis of longitudinal data. Adjusting for correlation between observations is one reason that modern longitudinal data analysis techniques are more appropriate than some previous methods of analyses.

For certain repeated measures designs, especially when the repeated measures are collected close in time or correlations among the repeated measures do not decay quickly, random effects alone may not adequately account for the dependency due to the repeated measures and a more complex covariance structure at level one may be needed [27].

2.2.3. Joint Modeling Approaches
In most longitudinal experiments, the number of outcomes measured repeatedly in the participating subjects exceeds one. Often, subject-matter research questions can be answered by

analyzing all outcomes separately. However, whenever interest is a comparison of longitudinal trends between outcomes, or interest is in the association between the outcomes and how that association evolves over time, joint analysis of all outcomes is required [29]. Joint modeling was also used in the context of jointly studying time to clinical event and repeated measures on surrogate outcomes [30]. Others included joint modeling of the multilevel item response theory (MLIRT) and Cox's proportional hazard model for time to dependent terminal event with shared random effects to link the two models [31]. A modeling framework for MLIRT also referred to as latent regression was widely considered and was based on the idea that the observed measurements are a result of some imperfect interaction between subject-specific latent variables and measurement- specific parameters [32, 33]. The latent traits are considered as response variables and are regressed on a set of covariates, hence the name of latent regression. The advantage of the MLIRT models resides in the separation of the measurement-specific parameters and subject-specific covariates from manifest data and simultaneous estimation of these parameters and covariate effects [34, 35]

A joint multivariate normal distribution was considered for the corresponding latent variables and each outcome was analyzed with a marginal dose-response model. The covariance matrix takes into account the correlation between outcomes and the correlation due to clustering. That was an important improvement of *Catalano and Ryan; Fitzmaurice and Laird* [36, 37] as model estimates of the correlation between outcomes and evolution of these correlations with dose were available.

Thiebauta et al. [38] studied the joint random effect model between the evolution of CD4 and HIV RNA. They reported that the bivariate random effects model was significantly better than two separate univariate random effects models with (p-value<0.0001). The other joint mixed effects model was studied by *Ferrari and Cribari-Neto* [39] on the evolution of occurrence and prevalence of antimicrobial resistant zoonotic agents. They used the beta-regression model and reported the correlation was estimated to be 0.95(p_value<0.0001). Their report shows strong positive and significant correlation between percentage resistant and prevalence.

Njagi et al. [40] jointly model the risk of re-hospitalization and the mean number of times a patient's HR measurements which were classified as "abnormal", with LVEF as a baseline covariate for chronic HF data. He first dichotomized HR into "normal" (50-90; coded 0) and "abnormal" (values higher than 90; coded 1) and the HR values less than 50 were not considered in the analysis. The baseline LVEF was considered as a covariate that indicates the fraction of blood being pumped out of the ventricle with each contraction. The test for a joint effect of ejection status on both processes was not statistically significant with (p_value=0.1650). Finally, they compared the results from the extended and the conventional model. Based on an AIC, they observed that their extended model provided an improvement to model fit, without compromising parsimony.

Bo and Sheng [41] proposed a joint modeling framework to jointly analyze the multivariate longitudinal data subject to dependent terminal events using the MLIRT sub-model and the Cox proportional hazard sub model. They link the two sub models together via shared random effects representing the subject-specific baseline disease severity and disease progression rate, respectively. They reported that the proposed joint model has a better fit than the reduced model in the analysis of the DATATOP data set.

The two joint models of HR with SBP and DBP were fitted by *Lambert and Vandenhende* [42]. They reported that there was a significant positive association between HR and DBP but not between HR and SBP. The effect of the sex was not significant on HR and DBP. Again they fitted the joint model for SBP and DBP and they said there was a significant positive association. As they reported that, the marginal mean HR was a significantly smaller for men than for women.

The other joint model study was the effects of training after discharge on readmission and re-hospitalization of patients with heart failure (randomized single-blind clinical trial) [43]. They compared the univariate and multivariate general linear mixed model results using the likelihood ratio test. According to their report, the random intercepts between cases were significant, expressing considerable variation at base line. The correlation between random intercepts in the model was statistically significant (correlation=0.905, p-value=<0.0001). They reported that; the multivariate model is better. In addition *Bediru* [44] reported in his study on the comparison and

computational survey of various univariate and multivariate learning curve models that, the bivariate model provided a slightly better fit than the univariate model. Santos et al. [45] also concluded that the multivariate approach performs better than the univariate approach in his study of comparing multivariate and univariate GARCH models to forecast portfolio value at risk.

A Joint Model for a Longitudinal Pulse Rate and Respiratory Rate of Congestive Heart Failure Patients was conducted by *Fissuh and Muleta* [7]. The separated and joint model was fitted with a random effect and independent measurement errors. A joint model with unstructured variance-covariance for random effect was a better fit for the data. He stated that quadratic fixed effects and random effects did not improve the model. The random effect model with linear time and intercept was selected. The baseline mean and standard deviation of Pr and Rr were 126.11 and 18.98 and 31.64 and 10.99 respectively. The association of the evolution for Pr and Rr was estimated to be (ρ=0.7054) which is statistically significant with 95% CI of (0.642, 0.769). Pr and Rr showed a decreasing pattern over time in both joint and separate models. The positive and significant associations were observed between the two end points and all covariates except LVEF and time.

Verbeke et al. (29) stated as; the idea of using random effects to account for the correlation between measurements within a subject can also are exploited to construct joint models for multivariate longitudinal outcomes. More specifically, it will be assumed that, conditionally on the random vectors b_k, Y_k follows a distribution with density $f(Y_k|b_k)$ possibly depending on the additional population-specific parameters θ_k, suppressed from notation. Some models assume all b_k to be identical, leading to so-called shared parameter models. Other models allow the different outcomes Y_k to be modeled with separate but correlated random vectors b_k, resulting in so-called random effects models.

3. Methodology

3.1. Study design

The study design was a longitudinal retrospective cohort study. A cohort of congestive heart failure patients were followed from December 2015 to January 2018 in Wachemo University Nigist Ellen Mohammed memorial Hospital.

3.2. Inclusion and exclusion criteria

Including congestive heart failure patients in the study would be whose age is 18 and above years and, who have followed minimum three and maximum nineteen visits from Dec 2015 to Jan 2018 in Wachemo University Nigist Ellen Mohammed memorial Hospital. The patients out of this criterion are not included in the study.

3.3. Target population

The target population is the study population in which the final results were applied. The results of this study (a final model which relative to another fit adequately) were applied for congestive heart failure patients. In this study all the congestive heart failure patients are the target population.

3.4. Data Source and Its Description

The source of data for this study is secondary data from retrospective cohort follow up of all congestive heart failure patients who have followed from December 2015 to January 2018 in Wachemo University Nigist Ellen Mohammed memorial Hospital.

The data was extracted from the follow-up patients chart. This chart was recorded by assigning an identification number per individual and contains epidemiological, laboratory and clinical information of all congestive heart failure patients. The data consists of 154 individuals with a minimum of three and maximum of nineteen measurements of pulse rate, respiratory rate, weight and other covariates measured per individual of adult congestive heart failure patients.

3.5. Variables

3.5.1. Dependent Variables

Three outcome variables are considered in this study; respiratory rate, heart rate and weight of Congestive heart failure patients for each individual measured at least three times.

3.5.2. Independent variables

Covariates (Independent variables): Age, Sex, Time, Place of Residence, NYHAC, Diagnostic History, LVEF (Left ventricle ejection fraction), valvular heart disease ,smoking status, obesity status, diabetes, diastolic blood pressure and systolic blood pressure.

3.6. Statistical Methods of Data Analysis

3.6.1. Exploratory Data Analysis

The first step in any model building process is exploratory data analysis. Data exploration is a very helpful tool in the selection of appropriate models to visualize the patterns of data relative to research interests. Analyses of longitudinal data compare profiles over time and, indeed, time might be viewed as the primary systematic effect to be investigated. The aim of this process is to understand the data structure and determine the relevant modeling approaches suitable for it. Thus, this study assessed the nature of the data by exploring individual profiles of respiratory rate, heart rate and weight of congestive heart failure patients, and the average evolution.

3.6.1.1. Exploring the Individual Profile

The individual profile plot was explored to show whether there is a noticeable pattern common to the most subjects in respiratory rate, heart rate and weight of congestive heart failure patients over follow-up time. These individual profiles provided some information on within and between- subject variability. It was used to identify general trends within subjects and may detect change over time that provides information about the variability at given times.

3.6.1.2. Exploring the Mean Structure

Examining the data for clues about the likely nature of the mean structure, to see how the mean profile changes over time, is essential for specifying the functional form of the mean response of the model. So as to understand the possible relationships among means over time, for balanced data, graphical inspection can be used by connecting the average values computed at each time point separately. If the data is not balanced and unequally spaced interval loess smoothing can be used instead. In this study, the loess smoothing technique was used and it gives an idea as for how the mean profile evolves over time. The results of this exploration were used in order to choose a fixed-effects structure for the linear mixed model.

3.6.2. Models for longitudinal data

The model which used in this study was a linear mixed effect model which is one of the mixed effect models. Mixed-effects model is a flexible and widely used model for the analysis of continuous longitudinal data to model the between and within subjects variation in the data. It has been a popular method to handle both balanced and unbalanced scenarios; and allows the inclusion of covariates. When both the fixed and the random effects contribute linearly to the response, the model is called a linear mixed-effects model.

3.6.2.1. Separate linear mixed model

Linear mixed model is a generalization of the standard linear model. It provides the flexibility of modeling not only the means of your data but also their variances and covariance. It is also an extension of the Linear Model that allows for incorporation of random effects and is represented in its most general fashions by *Molenberghs and Verbeke* [9].In this study there are three response variables, so the researcher has three separated models. The model for each response can be written as:

$$Y_i(t) = X_i(t)^T \beta + Z_i(t)^T b_i + \varepsilon_i(t) \text{ -- (1)}$$

Where, $Y_i(t)$: Measurement of univariate response in i^{th} patient at time t

$X_i(t)$: Vector of fixed covariate for i^{th} subject at time t (of dimension k)

$Z_i(t)$: Vector of random covariate for i^{th} subject at time t (of dimension q)

β : Vector of unknown parameters associated with a fixed covariate (of dimension k)

b_i: Vector of unknown parameters associated with random covariate for i^{th} subject (of dimension q), b_i ~ MVN (0, D) and $\varepsilon_i \sim \text{MVN}(0, R)$: Measurement error

ε_i is independent of b_i , where, X_i and Z_i are the fixed and random design of covariates, respectively, β is a vector of unknown fixed effects, b_i is a vector of unknown random effects and ε_i is the unknown measurement error. β Represents parameters that are the same for all subjects; b_i represents parameters that are allowed to vary over subjects. The assumptions for the models were checked.

3.6.2.1.1. Separated model estimation

Estimation of the parameters in LMM is usually based on maximum likelihood (ML) or restricted maximum likelihood (REML) estimation.

Maximum Likelihood Estimation: The ML method first maximizes the log-likelihood with respect to the variance parameters, while treating the fixed-effects parameters, β, as constant. Upon determining the variance parameter estimates, the fixed-effects parameters are then determined by finding the values of β which maximize the log likelihood, while treating the variance parameters as constant. The maximum likelihood estimates of variance components take no account of the degrees of freedom used in estimating fixed effects. This means that ML estimates of variance component have a downwards bias which increases with the number of fixed effects in the model.

Restricted Maximum Likelihood Estimation: This is another method that used to maximize the log-likelihood function. For this approach, the fixed-effects parameters, β, are eliminated from the log-likelihood equation, such that it will only be defined in terms of the variance parameters. The variance parameters are now estimated by maximizing the REML log-likelihood with regards to the variance parameters and values of β are found by maximizing the REML log likelihood with regards to the fixed-effects parameters, while treating the variance parameters as fixed. Given the nature of the REML likelihood, and its treatment of the fixed-effects as parameters, rather than as constants, the resulting variance parameter estimates are unbiased.

3.6.2.2. Joint Modeling

The joint model which investigated in this study were bivariate and multivariate linear mixed effect model which contain fixed, random effects and measurement errors and are referred to as β, bi, and ε_i respectively in the below equations of the joint models. The interdependency between the multiple outcomes and the longitudinal nature of the data were accounted for through the separate yet correlated random effects and measurement error. Fixed effects represent the average rate of change in the outcome attributed to specific covariates at a population level; however, the random effects represent the subject-specific rate of change. The correlation between repeated measures on a certain outcome pertaining to the same individual subject is accounted for through the measurement errors effects. The correlations between the different multiple outcomes are incorporated through the variance–covariance matrix of the random effects and correlation structure of measurement error.

The three end-points are longitudinally measured as a vector of responses, $Y_i(t)$ at each occasion with this model:

$$Y_i(t) = X_i(t)^T\beta + Z_i(t)^T b_i + \varepsilon_i(t)$$ --- (2)

15

$\varepsilon_i = [\varepsilon_i(t_1), \varepsilon_i(t_2), \varepsilon_i(t_3)]^T \sim \text{MVN}(0, R_i)$

$b_i \sim \text{MVN}(0, D), \quad \text{Cov}(b_i, \varepsilon_i) = 0$

D is the variance-covariance matrix of random effects.

$R_i = I_{n_i} \otimes \sum_{3X3}$ is the variance covariance matrix of 3 endpoints (symptoms) conditional on bi

Let $Y_i = \begin{bmatrix} Y_{1i}(t) \\ Y_{2i}(t) \\ Y_{3i}(t) \end{bmatrix}$, the response vectors for the subject i, with Y_{ki} the n_{ki} vector of the end points

k (k=1, 2, 3) with $n_{1i} = n_{2i} = n_{3i} = n_i$ so model for multivariate longitudinal Gaussian data is:

$$\left.\begin{aligned}
Y_{1i}(t) &= \mu_1(t) + \alpha_{1i} + b_{1i}(t) + \varepsilon_{1i}(t) \\
Y_{2i}(t) &= \mu_2(t) + \alpha_{2i} + b_{2i}(t) + \varepsilon_{2i}(t) \\
Y_{3i}(t) &= \mu_3(t) + \alpha_{3i} + b_{3i}(t) + \varepsilon_{3i}(t)
\end{aligned}\right\} \text{----------------------------------- (3)}$$

The number of parameters to be estimated is different for a model with random intercept and with random intercept and slope.

Random intercept model

It assumes the random effect is intercept. It is described as:

$$\left.\begin{aligned}
Y_{1i}(t) &= \mu_1(t) + \alpha_{1i} + \varepsilon_{1i}(t) \\
Y_{2i}(t) &= \mu_2(t) + \alpha_{2i} + \varepsilon_{2i}(t) \\
Y_{3i}(t) &= \mu_3(t) + \alpha_{3i} + \varepsilon_{3i}(t)
\end{aligned}\right\} \text{----------------------------------- (4)}$$

, $b_i = \begin{pmatrix} \alpha_{1i} \\ \alpha_{2i} \\ \alpha_{3i} \end{pmatrix} \sim N(0, D_{3x3})$ Where, D is the variance covariance between intercepts (covariance matrix for random effects). The number of parameters to be estimated is 6 attributed to the 3 variances and 3covariance between the intercepts.

$(\varepsilon_{1i}(t), \varepsilon_{2i}(t), \varepsilon_{3i}(t)) \sim N(0, \sum_{3X3})$, for all t Where $\sum_{3X3}$ is the variance covariance matrix of 3 endpoints (symptoms) conditional on b_i (covariance matrix for error components)

The three responses are joined using joint distributions of random intercept and measurement errors.

The random intercept and slope model

It assumes the random effect is intercepts and slope. It is described as:

$$\left.\begin{aligned}
Y_{1i}(t) &= \mu_1(t) + \alpha_{1i} + b_{1i}(t) + \varepsilon_{1i}(t) \\
Y_{2i}(t) &= \mu_2(t) + \alpha_{2i} + b_{2i}(t) + \varepsilon_{2i}(t) \\
Y_{3i}(t) &= \mu_3(t) + \alpha_{3i} + b_{3i}(t) + \varepsilon_{3i}(t)
\end{aligned}\right\} \text{----------------------------------- (5)}$$

$$, b_i = \begin{pmatrix} \alpha_i \\ b_i \end{pmatrix} = \begin{pmatrix} \begin{pmatrix} \alpha_{1i} \\ \alpha_{2i} \\ \alpha_{3i} \end{pmatrix} \\ \begin{pmatrix} b_{1i} \\ b_{2i} \\ b_{3i} \end{pmatrix} \end{pmatrix}, \sim N(0, D_{6x6}),$$ where D is the variance–covariance matrix for the

3 intercepts and 3 slopes associated with every outcome (covariance matrix for random effects) resulting in an increase of the number of parameters to be estimated.

(ε_{1i} (t), ε_{2i} (t), ε_{3i}(t)) $\sim$ N (0, $\sum_{3X3}$), for all t Where $\sum_{3X3}$ is the variance covariance matrix of 3 endpoints (symptoms) conditional on, b_i (covariance matrix for error components)

The three responses are joined using joint distributions of random intercept and slope and measurement errors.

There were three response variables and two random effects (random slope and intercept) for Pulse rate and respiratory rate and only random slope for weight, then there were 5 random effects. If researcher assume that random effects follow MVN(**0**, **D**) then **D** have $\binom{5}{2}$+5 covariance parameters and R have covariance $\binom{3}{2}$+3 unknown parameters. Therefore, together **D** and **R** have $\binom{5}{2}$+$\binom{3}{2}$+ 8 covariance parameters. The assumptions for the models were checked.

3.6.2.2.1. Joint Model Estimation

Gaussian quadrature

The Gaussian Quadrature approximates the integral of a function, with respect to a given kernel, by a weighted sum over predefined abscissas for the random effects. Unlike other numerical integration techniques, the abscissas are spaced unevenly throughout the interval of integration. With a modest number of Quadrature points, along with appropriate centering and scaling of the abscissas, the Gaussian Quadrature approximation can be highly effective see *Abramowitz and Stegun* for details [46]. *Pinheiro and Bates* [47] also suggested that, in the particular context of random-effects models, so-called adaptive Quadrature rules, where the numerical integration is centered on the estimates of the random effects, and the number of Quadrature points is then selected in terms of the desired accuracy. To illustrate the main ideas, they consider Gaussian

and adaptive Gaussian Quadrature, designed for the approximation of integrals of the form

$$\int f(z)\, \phi(z) dz$$ --- (6)

, for a known function *f(z)* and for $\phi(z)$ the density of the multivariate standard normal distribution. Therefore first standardize the random effects such that they get the identity covariance matrix. Then, the likelihood contribution for subject i equals

$$f(\text{yi}|\beta, G, \phi) = \int \prod_{j=1}^{n_i} f_{ij}(y_{ij}, b_i, \beta, \phi) f(b_i|G) db_i$$ ----------------------- (7)

, where b_i is q×1 dimensional vector of unknown random effects, bi ~ N (0, G)

β is a vector of fixed-effects parameters and ϕ is a vector containing the variance parameters

f (z) and for $\phi(z)$ denotes the density of the multivariate standard normal distribution

3.6.3. Correlation Structures

In longitudinal data analysis, when subjects are followed over time, there is a natural ordering of the data for each subject. Correlation structure is used to model dependence among observations, in the mixed-effect model; it is used to model dependency among the within-group errors [47]. The correlation between two within-group errors$\varepsilon_{ij}, \varepsilon_{ij'}$ is assumed to depend on some distance between them, and ρ is a vector of correlation parameters. The serial correlation structures in linear mixed-effects models are used to model dependency in the data observed sequentially over time and indexed by a one dimensional time vector [48]. The general serial correlation model is defined as

$$\text{Cor } (\varepsilon_{ij}, \varepsilon_{ij'}) = h(\rho)$$-- (8)

, Where h(.)-indicates autocorrelation function. Some of the most common serial correlations structures used in practice are:

Compound symmetry:-It is the simplest serial correlation structure, which assumes equal correlation among all within-group errors of the same subject. The corresponding correlation model is:

$$\text{Cor } (\varepsilon_{ij}, \varepsilon_{ij'}) = \rho$$ --- (9)

, this correlation model tends to be too simplistic for practical application.

 General (Unstructured):-The general correlation structure represents the other extreme in complexity to the compound symmetry structure. Each correlation is shown by a different parameter, the correlation function is h(ρ) = ρ^k; k = 1,2,. .. .While the general correlation model

tends to over parameterized model. It is useful for a few observations per subject that leads to a precise correlation with observations.

Autoregressive (AR):- *Box et al.* [49] described the family of correlation structure which includes different classes of linear stationary models: autoregressive models, moving average models, and a mixture of autoregressive-moving average models. Autoregressive models express the current observation as a linear function of previous observation plus a homoscedastic noise term. Let ε_t indexes an observation was taken at time t, μ_t indexes a noise term with $E[\mu_t] = 0$, and assumed independent of the previous observations.

$$\varepsilon_t = \phi_1 \varepsilon_{t-1} + \ldots \ldots + \phi_p \varepsilon_{t-p} + \mu_t, \ |\phi| < 1 \text{ ------------------------------------- (11)}$$

p is called the order of the autoregressive model, which denoted by AR(p). There are p correlation parameters in an AR (p) model, given by $\phi = (\phi 1, \phi 2, \ldots \ldots, \phi p)$. The AR (1) model is the simplest and one of the most useful autoregressive model. Its correlation function is

$$h(k; \phi) = \phi^k, \ k = 0; \ 1, \ldots \ldots$$

In the First-order autoregressive structure, it is assumed that the correlation between time points decrease as the distances in time increase, this implies that the number of time intervals between pairs of observation increases, the correlation decreases and approaches to zero. Measurements that are closer in time have higher correlation than measurements with longer time between them. This structure will often be more realistic than the compound symmetry and has the same number of parameters which often makes it more preferable.

3.6.4. Variable selection technique

To select significant variables, first, the main effect and main effect by time interaction were incorporated in to the initial candidate model. After that, avoid non-significant variables one by one starting from the most non-significant term which is called backward variable selection technique [47].

3.6.5. Model Comparisons or Selection Techniques

Model selection technique is one of the most frequently encountered problems in data analysis. In most observational epidemiological studies, investigators frequently attempt to construct the most desirable statistical model using the popular methods of forwarding, backward, and stepwise regression [47]. Of course, knowledge of the subject matter plays an important role in model selection, but is based strictly on the data; model selection is often carried out using one of

the automated procedures built into the software, of which the most popular method is perhaps stepwise model selection. These methods pose the problem of the arbitrary selection of the significance levels in allowing a variable to enter in to or to be dropped from the model during the selection process [50]. There is also the problem of multiple testing that comes with fitting and refitting the model. The issue is made more complicated in the case of repeated or longitudinal data were selecting the best model means not only to select the best mean structure but also the most optimal variance-covariance structure for model selection criteria, like AIC, BIC and likelihood ratio test used [51]. In this study, the most commonly known model selection criteria which are Akaike Information Criterion (AIC) [52], the Bayesian Information Criterion (BIC) [53] and Log-likelihood ratio test were used to select the best model.

$$AIC = -2\log L + 2p \qquad BIC = -2\log\text{Likelihood} + n\,Pr\log(N),$$

Where; -2 logL is twice the negative log-likelihood value for the model

P: - is the number of estimated parameters.

n pr: -denotes the total number of parameters in the model

N: - is the total number of observations used to fit the model. Smaller values of AIC and BIC reflect an overall better fit.

4. Results and discussion

4.1. Baseline Information and Descriptive Statistics

This study contains socio-demographic and clinical data of 154 patients at baseline whose age is 18 years and above receiving preferable drugs to improve the symptoms of CHF from December 2015 to January 2018 in Wachemo University Nigist Ellen Mohammed memorial Hospital. The three symptoms of congestive heart failure which used in this study are Pr, Rr, and Wh. The total number of visits from 154 subjects in the CHF treatments was 1346 and the number of visits per subject varied from 3 to 19 months with a mean and standard deviation of follow-up time 5.62 and 3.70 months respectively.

Table 10 of the study indicates the decreasing sample size over time due to deaths, dropouts, missed clinic visits and transferring to other hospital and also there is admitting and readmitting of the patients. Which is the frequency distribution of the responses at time t, that indicates the number of congestive heart failure patients possess each value of a response at a specific time.

Table 1: Number of CHF patients at baseline for categorical variables

Characteristics		Frequency	percent
Sex	Male	57	37.01
	Female	97	62.99
Place of residence	Urban	59	38.31
	Rural	95	61.69
NYHA	Class I	28	18.18
	Class II	32	20.78
	Class III	43	27.92
	Class IV	51	33.12
Diagnosis History	Severe anemia	55	35.71
	CHD	50	32.47
	ACF	10	6.49
	Others	39	25.33
Valvular heart disease status	Yes	101	65.58
	No	53	34.42

According to **table 1**, more than half 97 (62.99%) of the congestive heart failure patients are females and 57 (37.01%) are males on their sex; based on the place of residence more than half 95 (61.69%) of them are from rural and 59 (38.31%) are from urban. According to the New York

heart association classification congestive heart failure patients classified in to four classes. In this study most of the patients; 51(33.12%) of them are in class IV, 43(27.92%) are in class III, 32(20.78%) are in class II and 28(18.18%) are in class I. Based on the diagnostic history the patients in this study classified in to four groups which are severe anemia, CHD,ACF, and others. Among them most of the patients 55(35.71%) had severe anemia, 50(32.47%) had CHD, 10(6.49%) had ACF and 39(25.33%) had another diagnostic history. Also, most of the patients 101(65.58%) in this study had a valvular heart disease (Table 2).

The second (**Table:2)** contain the baseline statistics, according to it the baseline mean of Pr was 107.91 (SD= 26.53 beats per minutes) in female patients and 103.19 (SD=23.19 beats per minutes) in male patients, of Rr was 32.25 (SD= 11.73 breaths per minutes) in female patients and 30.32 (SD=10.92 breaths per minutes) in male patients and of Wh was 62.73 (SD= 9.32 breaths per minutes) in female patients and 67.98 (SD=10.65 breaths per minutes) in male patients. The baseline mean of Pr were 109.24 (SD= 25.19 beats per minutes) in urban resident patients and 104.25 (SD=25.42 beats per minutes) in rural resident patients, of Rr was 31.49 (SD= 11.09 breaths per minutes) in urban resident patients and 31.56 (SD=11.71 breaths per minutes) in rural resident patients and of Wh was 64.78 (SD= 11.41 breaths per minutes) in urban resident patients and 64.61 (SD=9.30 breaths per minutes) in rural resident patients.

From the New York heart association classes the baseline mean and standard deviation of Pr,Rr and weight of patients in class I were 94.64 (SD= 12.61 beats per minutes),25.14(SD=3.96 beats per minutes) and 61.50(SD=12.47 kg); in class II were 87.97 (SD= 18.54 beats per minutes),25.12(SD=4.50 beats per minutes) and 62.41(SD=9.08 kg); in class III were 101.86 (SD= 19.45beats per minutes),28.49(SD=7.54 beats per minutes) and 65.07(SD=9.18kg) and in class IV were 127.53 (SD= 24.18beats per minutes),41.63(SD=13.06beats per minutes) and 67.51(SD=9.52kg) respectively.

From the diagnostic history the baseline mean and standard deviation of PR,RR and weight of patients with severe anemia diagnostic history were 106.96 (SD= 27.38 beats per minutes),32.65(SD=12.44 beats per minutes) and 66.42(SD=9.35 kg); with CHD were 107.40 (SD= 26.25 beats per minutes),31.94(SD=12.05 beats per minutes) and 63.62(SD=11.07 kg); with ACF were 114.40 (SD= 21.28 beats per minutes),35.40(SD=12.79 beats per minutes) and

59.90(SD=11.54 kg) and with Others were 101.33 (SD= 22.04 beats per minutes),28.44(SD=8.09 beats per minutes) and 64.79(SD=9.33 kg) respectively. Finally, for the patients with a Valvular heart disease the baseline mean and standard deviation of Pr, Rr, and Wh were 111.65(SD=26.26 beats per minutes),34.83(SD=12.24 beats per minutes)and 66.07(SD=8.96kg)respectively.

Table 2: Baseline mean and standard deviations of Pr, Rr, and Wh at each characteristics.

Characteristics		Pulse Rate		Respiratory Rate		Weight	
		Mean	STD	Mean	STD	Mean	STD
Sex	Male	103.19	23.19	30.32	10.92	67.98	10.65
	Female	107.91	26.53	32.25	11.73	62.73	9.32
Place of residence	Urban	109.24	25.19	31.49	11.09	64.78	11.41
	Rural	104.25	25.42	31.56	11.71	64.61	9.30
NYHA	Class I	94.64	12.61	25.14	3.96	61.50	12.47
	Class II	87.97	18.54	25.12	4.50	62.41	9.08
	Class III	101.86	19.45	28.49	7.54	65.07	9.18
	Class IV	127.53	24.18	41.63	13.06	67.51	9.52
Diagnosis History	Severe anemia	106.96	27.38	32.65	12.44	66.42	9.35
	CHD	107.40	26.25	31.94	12.05	63.62	11.07
	ACF	114.40	21.28	35.40	12.79	59.90	11.54
	Others	101.33	22.04	28.44	8.09	64.79	9.33
Valvular heart disease status	Yes	112.65	26.26	34.83	12.24	66.07	8.96
	No	93.79	18.16	25.24	5.95	62.02	11.68

According to the **table 3,** the baseline mean and standard deviation of Pr were 106.16 and 25.37 beats per minutes, of the Rr, were 31.53 and 11.44 breaths per minutes, of the Wh, were 64.68 and 10.12 breaths per minutes, of the age, were 49.29 and 16.15 breaths per minutes, of the systolic blood pressure, were 121.90 and 23.43 breaths per minutes and of the diastolic blood pressure were 75.53 and 14.58 breaths per minutes

Table 3: Baseline mean and standard deviation for continuous variable

Variables	Mean	STD	Minimum	Maximum
Pulse rate	106.16	25.37	40	172
Respiratory rate	31.53	11.44	18	62
Weight	64.68	10.12	40	86
Age	49.29	16.15	18	85
Systolic blood pressure	121.90	23.43	79	175
Diastolic blood pressure	75.53	14.58	43	122
Left ventricular ejection fraction	34.37	13.52	19	82

4.2 Explanatory data analysis

4.2.1 Individual Profile Structure

The individual profile plot helps to identify the general trend within subjects over time and to decide which random effects to include in a model and what is the covariance structure this random effects should have.

Figure 1.A indicates that, the variation of pulse rate at starting time was higher than that of the end time. There is a high variation between congestive heart failure patients at the beginning time of the follow-up than at the end. The profile also shows a linear change of pulse rate over time. Some line`s on the plot show increasing trend. The graph suggests in generally a negative linear relationship between pulse rate and follow-up time. The value of pulse rate for some congestive heart failure patient changes non-linearly over time with a different intercept and slope, which are both random. The graph again indicates that there is much variability in pulse rate between and within congestive heart failure patients.

According to **Figure 1.B**, the variation of respiratory rate at starting time was higher than that of the end time. There is a high variation between congestive heart failure patients at the beginning time of the follow-up than at the end. The profile also shows a linear change of respiratory rate over time. Some lines on the plot show an increasing trend. The graph suggests a negative linear relationship between respiratory rate and follow-up time. The value of respiratory

rate for some congestive heart failure patient changes non-linearly over time with a different intercept and slope, which are both random. The graph again indicates that there is much variability in respiratory rate between and within congestive heart failure patients.

The below **figure 1.C** indicates that; the variation of weight at starting time was higher than that of the end time. There is a high variation between congestive heart failure patients at the beginning time of the follow-up than at the end. The profile also shows a linear change of weight over time. Some lines on the plot show an increasing trend. The graph suggests a negative linear relationship between weight and follow-up time. The value of weight for each congestive heart failure patient changes linearly over time with a different intercept, which is random. The graph again indicates that there is considerable variability in weight between and within congestive heart failure patients.
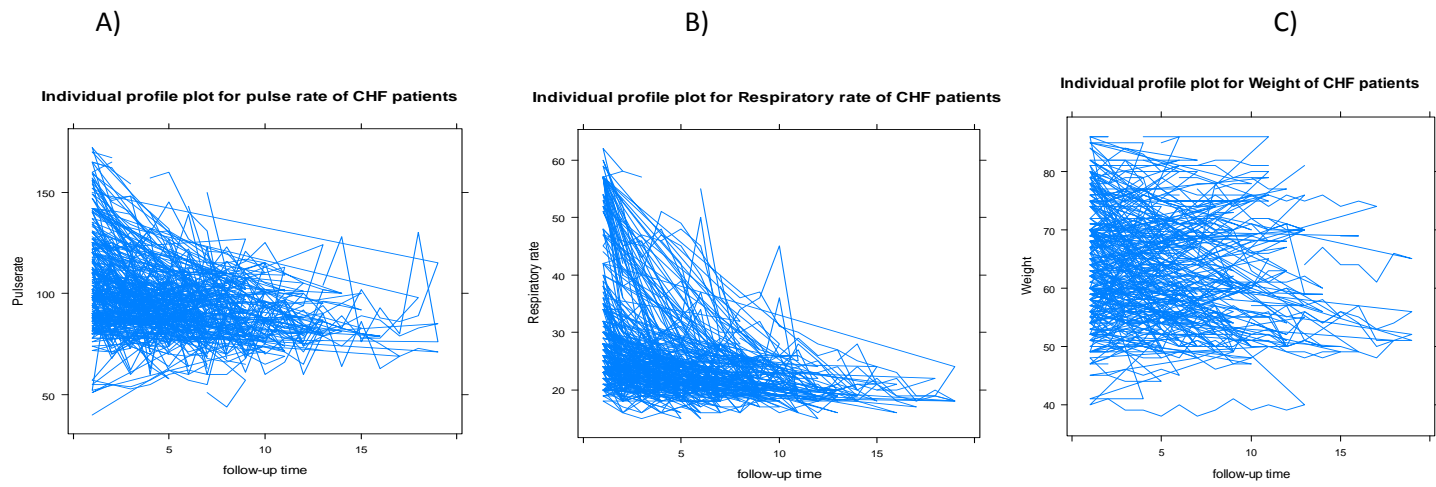
A)

B)

C)

Figure 1: Individual profile plot for Pr, Rr and Wh of CHF patients

A)

B)

C)
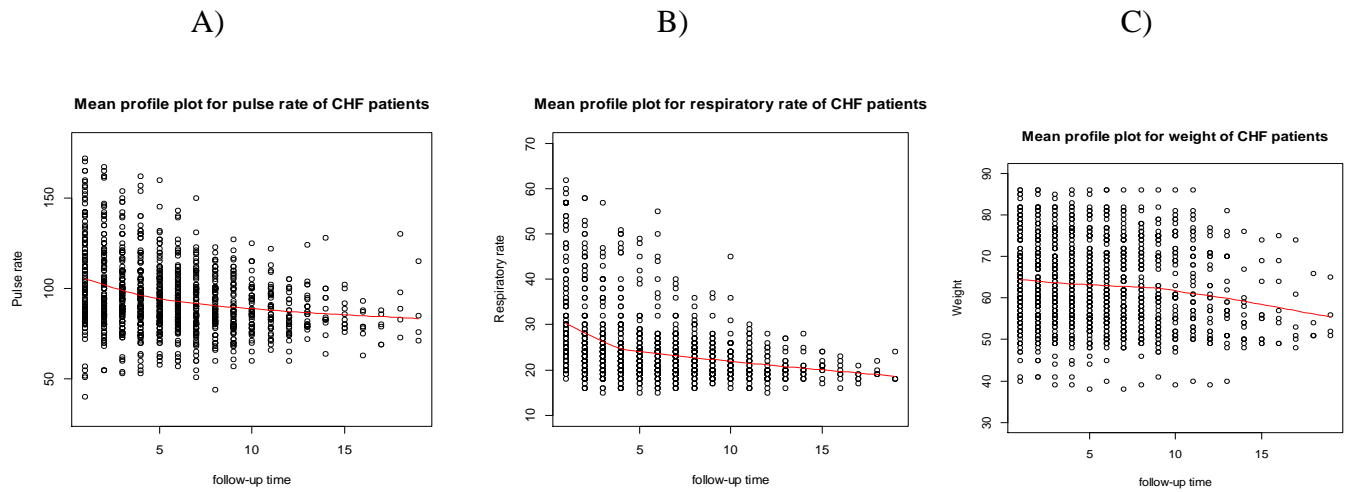
Figure 2: Mean profile plot for P r, Rr, and Wh of CHF patients

4.2.2 Mean structure

The mean profile describes how the profile for a number of relevant sub population evolves over time and also useful to choose a fixed-effects structure for linear mixed model structure. A loess smoothing technique was used which is the appropriate method for the data with missing value.

According to **Figure 2.A**, the average pulse rate of congestive heart failure patient not shows a clear linear trend with follow-up time. It seems upward parabola. The plots show that negative linear pattern between pulse rate and follow-up time of the patient. This means that; as the follow-up time of patients become to increase the average progression of pulse rate decline.

Figure 2.B indicates that; the average respiratory rate of congestive heart failure patient does not show a clear linear trend with follow-up time. It decreased with high rate up to the fourth month and then after there is a decreasing rate. It shows the negative pattern between respiratory rate and follow-up time of the patient. This means that; as the follow-up time of patients become to increase the average progression of respiratory rate decline.

The mean profile plot of weight in **Figure 2.C** indicates; the average weight of congestive heart failure patient is shown a linear trend with follow-up time. The plot seems that negative linear pattern between weight and follow-up time of the patient. This means that; as the follow-up time of patients to increase the average progression of weight becomes decline. The direction of the line is the same to the mean profile plots line of pulse rate and respiratory rate but the rate is not the same.

The mean profile plot of pulse rate by gender of **Figure 3.A** indicates, the average pulse rate of the patient seems like non-linear patterns in the groups, especially for male patients. It shows the average pulse rate of female patients is higher than that of males and no interaction effects between sex and follow-up time. It shows also have a decreasing pattern over follow-up time. It seems that the variability is almost the same among the two groups.

According to the mean profile plot of the respiratory rate by gender in **Figure 3.B**, the average respiratory rate of the patient for gender seems like a non-linear pattern, especially for males. The average line for females higher than that of males and no interaction effects between

sex and follow-up time. It shows a decreasing pattern over follow-up time. It seems that the variability is almost the same among the two groups.

The mean profile plot of weight by gender in **Figure 3.C** indicates, the average weight of the patient for gender seems like a linear pattern. The average line for a male is higher than that of female and no interaction effects between sex and follow-up time. It shows a decreasing pattern over follow-up time. It seems that the variability is almost the same among the two groups.

The mean profile plot of pulse rate by New York Heart Association Class in **Figure 4.A** indicates, the average pulse rate of the patient for New York Heart Association Class of patients seems like non-linear patterns. The average pulse rate for class IV is higher than that of all class up to the follow-up time 15. It shows a decreasing pattern over follow-up time. The variability is not the same in the classes. There is an interaction effect between New York Heart Association Class and follow-up time.

The mean profile plot of the respiratory rate by New York Heart Association Class in **Figure 4.B** indicates that; the average respiratory rate for New York Heart Association Class of patients seems like a non-linear pattern with the mean for class IV is higher than that of all class up to the follow-up time 13. It shows a decreasing pattern over follow-up time. The variability is not the same in the classes. There is an interaction effect between New York Heart Association Class and follow-up time.

The mean profile plot of weight by New York Heart Association Class **Figure 4.C** indicates, the average weight for New York Heart Association Class of patients seems like a linear pattern with the mean profile for class IV is higher than that of all class up to the follow-up time 12. It shows a decreasing pattern over follow-up time. The variability is not the same in the classes. A graph also indicates that the interaction effect between New York Heart Association Class and follow-up time.

According to **Figure 5.A**, the mean profile for the residence of patients seems like a non-linear pattern especially in rural residents. The mean profile for an urban resident is higher than that of rural residents up to the follow-up time 11. It shows a decreasing pattern over follow-up

time. It indicates that the variability is not the same among the urban and rural residents. There is an interaction effect between place of the residence of patients and follow-up time.

From the mean profile plot of the respiratory rate by the of the congestive heart failure patients, we can observe that the mean profile for the residence of patients seems like a non-linear pattern with the mean profile for an urban resident is higher than that of rural residents up to the follow-up time 11. It shows a decreasing pattern over follow-up time. It indicates that the variability is almost the same among the urban and rural residents. There is an interaction effect between the place of residence of patients and follow-up time.

The mean profile plot of weight by the residence of the congestive heart failure patients indicates that, the average weight of patients for residence seems like a linear pattern with the almost the same mean profile for both residences. It shows a decreasing pattern over follow-up time. It indicates that the variability is almost the same among the urban and rural residents. There is no clear interaction effect between the place of residence of patients and follow-up time.

According to **Figure 6.A**, the average line for the diagnostic history of patients seems like a non-linear pattern. The mean for ACF diagnostic history is higher than that of all another diagnostic history up to the follow-up time 12. It shows a decreasing pattern over follow-up time. It seems that the variability is not the same among the diagnostic history. There is an interaction effect between diagnostic history and follow-up time.

The mean profile plot of the respiratory rate by diagnostic history indicates that, the average respiratory rate of patients for diagnostic history seems like a non-linear pattern with the mean profile for ACF diagnostic history is higher than that of all another diagnostic history up to the follow-up time 5. It shows a decreasing pattern over follow-up time. It seems that the variability is not the same among the diagnostic history. There is an interaction effect between diagnostic history and follow-up time.

The mean profile plot of weight by diagnostic history indicates that; the average weight for the diagnostic history of patients seems like a non-linear pattern. The average weight of patients with diagnostic history sever anemia is highest of all another diagnostic history up to the follow-up time 10. It shows a decreasing pattern over follow-up time for all another diagnostic history

except ACF. It seems that the variability is not the same among the diagnostic history. There is an interaction effect between diagnostic history and follow-up time.

4.3 Variance Function and Correlation

4.3.1 Exploring Variance Function

Figure 7.A show that there is no constant variance of the patient's pulse rate .At the baseline the variation of pulse rate is very high among the congestive heart failure patients and its slope is decreasing but not up to the end of the follow-up time, in the end, it is very high.

Figure 7.B shows that there is no constant variance of the patient's respiratory rate .At the baseline the variation of respiratory rate is very high among the congestive heart failure patients and its slope is decreasing up to the end of the follow-up time.

Figure 7.C shows that there is no constant variance of the patient's weight .At the baseline the variation of weight is high and it continues up to a follow-up time 10,at this point it is very high then it continues decreasing but it is not constant decreasing up to the end of follow-up time.

4.4. Modeling the outcomes
4.4.1 Separated model
4.4.1.1 Selecting fixed effects for separated model
To selects significant variables, the backward variable selection method was used. At the first time, a full model that includes all covariates (main effects) and interaction effects with time was fitted and removed the most insignificant variable. Then we fitted a reduced model, again we did the same as previews that removing the most insignificant variables and we did the same procedure until we get the candidate model. To fit the log of pulse rate and a log of respiratory rate models we used to intercept and linear time and for a log of weight intercept as a random effect with covariance structure compound symmetry and ML estimation method to select the fixed effects.

Finally a model with fixed effects sqrtime, Sex, Age, Residence, New York Heart Association Class, Systolic blood pressure, Diastolic blood pressure and New York Heart Association Class *sqrtime was selected as a final model for log of pulse rate with relatively small values of **AIC**= **-3788.9**, **BIC**= **-3737.2** and log-Likelihood ratio test with P-value of

<0.0001.For log of respiratory rate a model with fixed effects sqrtime, Sex, Age, New York Heart Association Class, Systolic blood pressure, Diastolic blood pressure, Diagnostic history, Valvular heart disease, New York Heart Association Class *sqrtime, was selected with relatively small values of **AIC= -3430.4, BIC= -3369.7** and log-Likelihood ratio test with P-value of <0.0001.A model with Time, Sex, Age, New York Heart Association Class, Systolic blood pressure, Diastolic blood pressure, Smoking New York Heart Association Class *Time and Diagnostic history*Time was selected for log of weight with relatively small values of **AIC= -6179.4, BIC= -6121.7** and log-Likelihood ratio test with P-value of <0.0001.

4.4.1.2 Selecting random effects for the separated model

After selecting the fixed effects, we need to select a set of random effects which can help in determining a model. In this section the aim is to select the random effect model of the rate of change of a log of pulse rate, a log of respiratory rate and log of weight measured over time including the selected fixed effects.

Thus, four different models with different random effects starting from the linear regression model (no random effects) have been explored and the summary of the models are in **table 4**. According to the information criteria of the models in the table the random effects model is better than that of the linear regression model. It also suggests that including the linear and quadratic time effect as random effects does not improve the model fit for all outcomes. When we compare a model with only intercept and model with intercept and linear time effects the linear time effect model fit better than only intercept model for a log of pulse rate and a respiratory rate that of the intercept model fits better for a log of weight.

Table 4: Random effects models

	The models	AIC	BIC	-2LL
For log of pulse rate	Without random effects	-3352.2	-3347.1	-3354.2
	With only intercept	-3663.9	-3654.8	-3669.9
	With intercept and slope	**-3678.3**	**-3669.2**	**-3684.3**
	Intercept, Linear and Quadratic Slope	-3604.2	-3595.1	-3610.2
For log of respiratory rate	Without random effects	-2691.6	-2686.5	-2693.6
	With only intercept	-3204.1	-3195.0	-3210.1
	With intercept and slope	**-3319.0**	**-3309.9**	**-3325.0**
	Intercept, Linear and Quadratic Slope	-3149.8	-3140.7	-3155.8
For log of weight	Without random effects	-3339.5	-3334.4	-3341.5
	With only intercept	**-6014.5**	**-6005.4**	**-6020.5**
	With intercept and slope	-5676.7	-5667.6	-5682.7
	Intercept, Linear and Quadratic Slope	-3928.8	-3922.7	-3932.8

4.3.1.3 Selecting variance-covariance structure of random effects for the separated model

According to **table 5**, the models for random effects with covariance structure unstructured and heterogeneous compound symmetry have the same value of information criteria for a log of pulse rate and a log of respiratory rate and the unstructured and Toeplitz covariance structure for a log of weight. The unstructured variance-covariance structure model was selected commonly for all three models.

Table 5: Variance-covariance structure of random effects

	Model	AIC	BIC	-2 res LL
For pulse rate	Compound symmetry	-3678.3	-3669.2	-3684.3
	Heterogeneous compound symmetry	**-3810.0**	**-3797.9**	**-3818.0**
	Unstructured	**-3810.0**	**-3797.9**	**-3818.0**
	Autoregressive	-3678.3	-3669.2	-3684.3
	Toeplitz	-3678.3	-3669.2	-3684.3
	Variance components	-3711.1	-3702.0	-3717.1
For respiratory rate	Compound symmetry	-3319.0	-3309.9	-3325.0
	Heterogeneous compound symmetry	**-3504.1**	**-3492.0**	**-3512.1**
	Unstructured	**-3504.1**	**-3492.0**	**-3512.1**
	Autoregressive	-3319.0	-3309.9	-3325.0
	Toeplitz	-3319.0	-3309.9	-3325.0
	Variance components	-3367.2	-3358.1	-3373.2
For weight	Compound symmetry	-6014.5	-6005.4	-6020.5
	Heterogeneous compound symmetry	-6014.5	-6005.4	-6020.5
	Unstructured	**-6016.5**	**-6010.5**	**-6020.5**
	Autoregressive	-6014.5	-6005.4	-6020.5
	Toeplitz	**-6016.5**	**-6010.5**	**-6020.5**
	Variance components	**-6016.5**	**-6010.5**	**-6020.5**

4.4.1.4. Selecting correlation structure of errors for separate model

According to **table 11**, the final model with autoregressive correlation structure was preferred for all Pr, Rr, and Wh model with respective small values of AIC, BIC and -2LL of **-3865.6**, **-3850.4** and **-3875.6**, AIC, BIC and -2LL of **-3578.9**, **-3563.7** and **-3588.9** and AIC, BIC and -2LL of **-6311.2, -6302.1**and **-6317.2** respectively. To identify this best correlation structure of error REML estimation method was used as random effects variance-covariance structure.

4.4.1.5 The results of final separated mixed effect model
The separated mixed effect models for the three symptoms of CHF syndrome Pr, Rr, and Wh was fitted by assuming there is no correlation between them. After the different procedures of the model building the final mixed effect model was selected. It was fitted with random intercept and slope for a log of pulse rate and a log of respiratory rate and only random intercept for a log of weight with the unstructured variance-covariance structure of random effects and autoregressive correlation structure of measurement error. According to the **table 6**, the time, age, sex, New York Heart Association Class, systolic blood pressure and diastolic blood pressure are common significant factor for all outcomes. The interaction effect of New York Heart Association Class and time for a log of pulse rate and a log of respiratory rate, Valvular heart disease for a log respiratory rate and Smoking and the interaction effect of Diagnostic history and time for a log weight are the significant factors in the separated model.

The estimated parameters for intercept of a log of Pulse rate, a log of Respiratory rate and a log of weight is 1.8902, 1.4401 and 1.6873 with standard error of 0.02265, 0.03207 and 0.01726 represents an average of log of Pulse rate, log Respiratory rate and log of weight during the first follow up time respectively and excluding all covariates in the model. Among all covariates, age, systolic blood pressure, and diastolic blood pressure were positively associated with all the three outcomes that mean the increase of age, systolic and diastolic blood pressure leads to a particular increase on the three outcomes. There was evidence that time had a negative effect (-0.04484 (0.001815), -0.1161(0.007169) and -0.00465(0.001130)) on the evolution of Pr, Rr, and Wh respectively. Sex was significantly associated with Pr, Rr, and Wh outcomes; male patients had -0.02184 (se=0.007002) points lower over evolution of Pr (P=0.0019), -0.03344 (se=0.01062) points lower over evolution of Rr (P=0.0017) and 0.03561(0.009957) points higher over evolution of Wh (P=0.0004) compared to females.

Similarly, NYHA class was significantly associated with all Pr , Rr, and Wh, thus, patients under NYHA class I had -0.1142 points lower over evolution of Pr (P<0.0001) and -0.2038 points lower over evolution of Rr (P<0.0001), class II had -0.1341 points lower over evolution of PR (P<0.0001) and -0.2028 points lower over evolution of Rr (P<0.0001) and -0.02164 points lower over evolution of Wh (P=0.0208) and class III had -0.07104 points lower

over evolution of Pr (P= 0.0002) and -0.1785 points lower over evolution of Rr (P<0.0001) relative to class VI.

Generally, in separated mixed effect model there is also decreasing pattern of Pr and Rr over time with respect to associated risk factors on respective treatments in the separate mixed model. Age had a positive effect on all outcomes. There is a significant difference between sex and New York Heart Association Class commonly for the three outcomes of congestive heart failure (Pr, Rr, and Wh) at 5% level of significance.

Table 6: *Parameter estimates and standard errors for the separate models of the Pr, Rr, and Wh*

Separated

Log of Pulse Rate

Effects	Estimate (SE)	P-value
Intercept ($\hat{\alpha}_{10}$)	1.8902 (0.02265)	<.0001
Time ($\hat{\alpha}_{11}$)	-0.04484 (0.001815)	<.0001
Age ($\hat{\alpha}_{13}$)	0.000968 (0.000219)	<.0001
Sex ($\hat{\alpha}_{14}$)	-0.02184 (0.007002)	0.0019
NYHAC I ($\hat{\alpha}_{15}$)	-0.1142 (0.02228)	<.0001
II ($\hat{\alpha}_{16}$)	-0.1341 (0.02140)	<.0001
III ($\hat{\alpha}_{17}$)	-0.07104 (0.01911)	0.0002
Sys ($\hat{\alpha}_{18}$)	0.001020 (0.000114)	<.0001
Dias ($\hat{\alpha}_{19}$)	0.000741 (0.000168)	<.0001
NYHAC*Time I*Ti ($\hat{\alpha}_{110}$)	0.03382 (0.008587)	<.0001
II*Ti ($\hat{\alpha}_{111}$)	0.02953 (0.008551)	0.0006
III*T ($\hat{\alpha}_{112}$)	0.01636 (0.007562)	0.0308
Sigma1 (σ_1)	0.002022 (0.000159)	<.0001
ε_1	0.3453 (0.04972)	<.0001
$\hat{\sigma}^2_{b10}$	0.004939 (0.001123)	<.0001
$\sigma_{b10,b11}$	-0.00159 (0.000427)	0.0002
$\hat{\sigma}^2_{b11}$	0.000670 (0.000181)	0.0001

Log of rate

Effects	Estimate (SE)	P-value
Intercept ($\hat{\alpha}_{20}$)	1.4401 (0.03207)	<.0001
Time ($\hat{\alpha}_{21}$)	-0.1161 (0.007169)	<.0001
Age ($\hat{\alpha}_{23}$)	0.000918 (0.000339)	0.0069
Sex ($\hat{\alpha}_{24}$)	-0.03344 (0.01062)	0.0017
NYHAC I ($\hat{\alpha}_{25}$)	-0.2038 (0.03356)	<.0001
II ($\hat{\alpha}_{26}$)	-0.2028 (0.03151)	<.0001
III ($\hat{\alpha}_{27}$)	-0.1785 (0.02646)	<.0001
Sys ($\hat{\alpha}_{28}$)	0.000921 (0.000123)	<.0001
Dias ($\hat{\alpha}_{29}$)	0.000647 (0.000179)	0.0003
Vhds ($\hat{\alpha}_{210}$)	0.03854 (0.01434)	0.0073
NYHAC*Time I*Ti ($\hat{\alpha}_{211}$)	0.07713 (0.01149)	<.0001
II*Ti ($\hat{\alpha}_{212}$)	0.06895 (0.01163)	<.0001
III*Ti ($\hat{\alpha}_{213}$)	0.05666 (0.009849)	<.0001
Sigma1 (σ_2)	0.002422 (0.000231)	<.0001
ε_2	0.4217 (0.05425)	<.0001
$\hat{\sigma}^2_{b20}$	0.01305 (0.002235)	<.0001
$\sigma_{b20,b21}$	-0.00418 (0.000832)	<.0001
$\hat{\sigma}^2_{b21}$	0.001745 (0.000351)	<.0001

Log of Weight

Effects	Estimate (SE)	P-value
Intercept ($\hat{\alpha}_{30}$)	1.6873 (0.01726)	<.0001
Time ($\hat{\alpha}_{31}$)	-0.00465 (0.001130)	<.0001
Age ($\hat{\alpha}_{31}$)	0.001406 (.000294)	<.0001
Sex ($\hat{\alpha}_{32}$)	0.03561 (0.009957)	0.0004
NYHAC I ($\hat{\alpha}_{33}$)	-0.00322 (0.008995)	0.7206
II ($\hat{\alpha}_{34}$)	-0.02164 (0.009347)	0.0208
III ($\hat{\alpha}_{35}$)	0.004438 (0.005844)	0.4478
Sys ($\hat{\alpha}_{36}$)	0.000191 (0.000034)	<.0001
Dias ($\hat{\alpha}_{37}$)	0.000198 (0.000050)	0.0003
Smok ($\hat{\alpha}_{38}$)	0.07656 (0.02494)	0.0022
DIG HIS*Time Other*Ti ($\hat{\alpha}_{39}$)	0.002734 (0.001202)	0.0231
Sev.*Ti ($\hat{\alpha}_{310}$)	0.001959 (0.00119)	0.0999
CHD*Ti ($\hat{\alpha}_{311}$)	0.003152 (0.001175)	0.0074
Sigma (σ_3)	0.000329 (0.000036)	<.0001
ε_3	0.6535 (0.03944)	<.0001
$\hat{\sigma}^2_{b30}$	0.003383 (0.00041)	<.0001

Where, SYs and Dias=Systolic and diastolic blood pressure, Ti =time, NYHAC=New York Heart Association Class, sev=sever anemia, CHD=Coronary heart disease, ACF=Acute coronary failure, Vhds=valvular heart disease, and smok=smoking status

4.4.2 Bivariate model

4.4.2.1 Selecting fixed effects for the bivariate model

To select significant variables for the bivariate model, backward variable selection method was used which is the same procedure as a separated model to select the fixed effects of the bivariate models. A model with fixed effects sqrtim1, sqrtim2, int1, int2, sex1, sex2, age1, age2, resi1, New York Heart Association C1, New York Heart Association C2,sys1, sys2, diast1, diast2, diag2, hyp1 hyp2 sqrtim1*NYHAC1 sqrtim2*NYHAC2 and sqrtim2*diag2 was selected as a final joint model of log of pulse rate and log of Respiratory rate with relatively small values of AIC= -7113.9, BIC= -7010.6 and log-Likelihood ratio test with P-value of <0.0001. For log of respiratory rate and log of weight a model with fixed effects sqrtim1, tim2, int1, int2, sex1, sex2, age1, age2, NYHAC1, sys1, sys2, diast1, diast2, diag1, smoks2, hyp1, vhds1 and sqrtim1*NYHAC1was selected with relatively small values of AIC= -8558.0, BIC= -8482.1 and log-Likelihood ratio test with P-value of <0.0001. A model with sqrtim1, tim2, int1, int2, sex1, sex2, age1, age2, resi1, NYHAC1, sys1, sys2, diast1, diast2, smoks2, hyp1, sqrtim1*NYHAC1 and tim2*diag2 was selected for a joint log of Pulse rate and log of weight with relatively small values of AIC= -8959.2, BIC= -8883.3 and log-Likelihood ratio test with P-value of <0.0001.

4.4.2.2 Selecting covariance structure of random effects for the bivariate models

According to **table 12**, the final bivariate models for random effects with unstructured covariance structure was preferred for all pairs (Pr, Rr), (Rr,Wh) and (Pr,Wh) with respective small values of AIC, BIC and -2LL of **-7403.4**, **-7370.0** and **-7425.4**, AIC, BIC and -2LL of **-8816.6**, **-8728.5** and **-8874.6** and AIC, BIC and -2LL of **-8977.4, -8956.2** and **-8991.4** respectively. To identify this best variance-covariance structure of random effects REML estimation method was used.

4.4.2.3 Selecting correlation structure of measurement errors for bivariate models

According to the **table 13**, the most use full correlation structures of error is Autoregressive for all bivariate model since a model with this correlation structure has the smallest AIC values, BIC values, and the Log Likelihood scores compared to the other models. To identify this best variance-covariance structure of error REML estimation method was used with the selected fixed

effects and random effects with unstructured variance-covariance structure model which was selected as a final best model.

4.4.2.4 The results of final bivariate mixed effect model

The three bivariate mixed effect models for the three symptoms of CHF syndrome Pr, Rr, and Wh was fitted with an unstructured covariance structure of random effects and autoregressive correlation structures of the measurement errors. To select the final best model the different procedures was passed as like the procedures passed on the separated models These models are the same as the separated models except the sets of random intercepts and slopes for each response are now correlated rather than independent. The final bivariate models (Pr and Rr), (Pr and Wh), and (Rr and Wh) with smallest value of AIC=-7578.6, BIC= -7536.1 and -2LL = -7606.6 for log of pulse rate and respiratory rate, AIC= -10258.1, BIC= -10227.8 and -2LL = -10278.1 for log of pulse rate and weight and AIC= -10181.6, BIC= -10093.5 and -2LL = -10239.6 for log of respiratory rate and weight was selected.

According to **table 7**, the fixed-effect intercept coefficient $\hat{\alpha}_{10}$= 1.9089(se= 0.02247) and $\hat{\alpha}_{20}$= 1.4674(se=0.03046) represents an estimate of the average log of Pr and log of Rr of the patients respectively at time=0 and excluding all covariates in the bivariate model log of Pr and log of Rr. In the same way, the fixed-effect intercept coefficient $\hat{\alpha}_{10}$= 1.8948(0.02259), $\hat{\alpha}_{30}$=1.6789(0.01570) represents an estimate of the average log of Pr and log of Wh of the patients respectively at time=0 and excluding all covariates in the bivariate model log of Pr and log of Wh, $\hat{\alpha}_{20}$= 1.4424 (0.03154), $\hat{\alpha}_{30}$=1.6790 (0.01549) represents an estimate of the average log of Rr and log of Wh of the patients respectively at time=0 and excluding all covariates in the bivariate model log of Rr and log of Wh. All the fixed effect parameters in the model are statistically significant except NYHACII for bivariate model log of Pr and log of Rr.

In addition, sex, sqrtime, age, NYHAC, systolic blood pressure, diastolic blood pressure, and NYHAC *time was significantly associated with both Pr and Rr outcomes in bivariate model Pr and Rr commonly and the all parameters are statistically significant. For a bivariate model Pr and Wh time, age, sex, systolic blood pressure, and diastolic blood pressure was significantly associated with both Pr and Wh outcomes commonly but the common parameters are not statistically significant. Generally, as it is indicated in the results in **table 7**,

both Pr and Rr have decreasing pattern throughout the follow up with respective clinical treatments.

Table 7: *Parameter estimates and standard errors for the bivariate models of the Pr, Rr, and Wh.*

Joint (Pr andRr)		Pulse rate(Pr)			**Respiratory rate(Rr)**	
Effects		Estimate(SE)	P-value	effects	Estimate(SE)	P-value
Inter($\hat{\alpha}_{10}$)		1.9089(0.02247)	<.0001	Inter($\hat{\alpha}_{20}$)	1.4674(0.03046)	<.0001
Time($\hat{\alpha}_{11}$)		-0.05208(0.005436)	<.0001	Time($\hat{\alpha}_{21}$)	-0.1180(0.007196)	<.0001
Age($\hat{\alpha}_{12}$)		0.000997(0.000221)	<.0001	Age($\hat{\alpha}_{22}$)	0.001157(0.000336)	0.0006
Sex($\hat{\alpha}_{13}$)		-0.02185(0.007057)	0.0020	Sex($\hat{\alpha}_{23}$)	-0.03396(0.01078)	0.0017
NYHAC	I($\hat{\alpha}_{14}$)	-0.1239(0.02246)	<.0001	I($\hat{\alpha}_{24}$)	-0.2351(0.03190)	<.0001
	II($\hat{\alpha}_{15}$)	-0.1409(0.02169)	<.0001	II($\hat{\alpha}_{25}$)	-0.2170(0.03060)	<.0001
	III($\hat{\alpha}_{16}$)	-0.08153(0.01887)	<.0001	III($\hat{\alpha}_{26}$)	-0.1845(0.02629)	<.0001
Sys($\hat{\alpha}_{17}$)		0.000981(0.000110)	<.0001	Sys($\hat{\alpha}_{27}$)	0.000911(0.000120)	<.0001
Dias($\hat{\alpha}_{18}$)		0.000693(0.000163)	<.0001	Dias($\hat{\alpha}_{28}$)	0.000645(0.000175)	0.0002
NYHAC*Ti	I*T($\hat{\alpha}_{19}$)	0.03838(0.008594)	<.0001			
	II*T($\hat{\alpha}_{110}$)	0.03451(0.008733)	<.0001	I*T($\hat{\alpha}_{29}$	0.08186(0.01151)	<.0001
	III*T($\hat{\alpha}_{111}$)	0.02232(0.007433)	0.0027	II*T($\hat{\alpha}_{210}$)	0.06576(0.01176)	<.0001
Sigma1(σ_1)		0.001911(0.000134)	<.0001	III*T($\hat{\alpha}_{211}$)	0.06012(0.009725)	<.0001
	ε_1	0.3179(0.04645)	<.0001	Sigma1(σ_2)	0.002310(0.000200)	<.0001
	$\hat{\sigma}^2{}_{b10}$	0.005767(0.001154)	<.0001	ε_2	0.4015(0.05129)	<.0001
	$\sigma_{b10,b11}$	-0.00198(0.000448)	<.0001	$\hat{\sigma}^2{}_{b20}$	0.01411(0.002278)	<.0001
	$\hat{\sigma}^2{}_{b11}$	0.000872(0.000193)	<.0001	$\sigma_{b20,b21}$	-0.00460(0.000850)	<.0001
				$\hat{\sigma}^2{}_{b21}$	0.001958(0.000360)	<.0001
Common parameters	$\sigma_{b10,b20}$				0.006891(0.001264)	<.0001
	$\sigma_{b10,b21}$				-0.00261(0.000499)	<.0001
	$\sigma_{b20,b11}$				-0.00237(0.000514)	<.0001
	$\sigma_{b11,b21}$				0.001114(0.000210)	<.0001

A)

Joint(Pr and Wh)

Log of pulse rate			Log of weight		
Effects	Estimate(SE)	P-value	effects	Estimate(SE)	P-value
Inter($\hat{\alpha}_{10}$)	1.8948(0.02259)	<.0001	Inter($\hat{\alpha}_{30}$)	1.6789(0.01570)	<.0001
Time($\hat{\alpha}_{11}$)	-0.04553(0.005390)	<.0001	Time($\hat{\alpha}_{31}$)	-0.00432(0.001127)	0.0001
Age($\hat{\alpha}_{12}$)	0.000963(0.000218)	<.0001	Age($\hat{\alpha}_{32}$)	0.001532(0.000285)	<.0001
Sex($\hat{\alpha}_{13}$)	-0.02167(0.006980)	0.0019	Sex($\hat{\alpha}_{33}$)	0.03368(0.009926)	0.0007
NYHA I($\hat{\alpha}_{14}$)	-0.1187(0.02241)	<.0001	Sys($\hat{\alpha}_{34}$)	0.000188(0.000034)	<.0001
NYHA II($\hat{\alpha}_{15}$)	-0.1365(0.02141)	<.0001	Dias($\hat{\alpha}_{35}$)	0.000191(0.000050)	0.0001
NYHA III($\hat{\alpha}_{16}$)	-0.07217(0.01900)	0.0001	Smok($\hat{\alpha}_{36}$)	0.07510(0.02464)	0.0023
Sys($\hat{\alpha}_{17}$)	0.001007(0.000113)	<.0001	Diag*T I*T($\hat{\alpha}_{37}$)	0.002462(0.001199)	0.0403
Dias($\hat{\alpha}_{18}$)	0.000723(0.000166)	<.0001	Diag*T II*T($\hat{\alpha}_{38}$)	0.001522(0.001183)	0.1983
NYHA C*Ti I*T($\hat{\alpha}_{19}$)	0.03553(0.008611)	<.0001	Diag*T III*T($\hat{\alpha}_{39}$)	0.002798(0.001171)	0.0170
NYHA C*Ti II*T($\hat{\alpha}_{110}$)	0.03055(0.008532)	0.0004	Sigma1(σ_3)	0.000331(0.000036)	<.0001
NYHA C*Ti III*T($\hat{\alpha}_{111}$)	0.01677(0.007523)	0.0259	ε_3	0.6533(0.03924)	<.0001
Sigma1(σ_1)	0.002030(0.000164)	<.0001	$\hat{\sigma}^2_{b30}$	0.003389(0.000409)	<.0001
ε_1	0.3513(0.05056)	<.0001			
$\hat{\sigma}^2_{b10}$	0.004952(0.001139)	<.0001			
$\sigma_{b10,b11}$	-0.00160(0.000434)	0.0002			
$\hat{\sigma}^2_{b11}$	0.000668(0.000185)	0.0001			
Common parameters	$\sigma_{b10,b30}$			-0.00033(0.000471)	0.4812
	$\sigma_{b30,b11}$			0.000173(0.000198)	0.3826

B)

Joint(Rr and Wh)

Log of respiratory rate			Log of weight		
Effects	Estimate(SE)	P-value	effects	Estimate(SE)	P-value
Inter($\hat{\alpha}_{20}$)	1.4424 (0.03154)	<.0001	Inter($\hat{\alpha}_{30}$)	1.6790 (0.01549)	<.0001
Time($\hat{\alpha}_{21}$)	-0.1163 (0.007063)	<.0001	Time($\hat{\alpha}_{31}$)	-0.00208(0.000233)	<.0001
Age($\hat{\alpha}_{22}$)	0.000900 (0.000328)	0.0062	Age($\hat{\alpha}_{32}$)	0.001505(0.000281)	<.0001
Sex($\hat{\alpha}_{23}$)	-0.03313 (0.01030)	0.0013	Sex($\hat{\alpha}_{33}$)	0.03485(0.009763)	0.0004
NYHA C I($\hat{\alpha}_{24}$)	-0.2056 (0.03335)	<.0001	Sys($\hat{\alpha}_{34}$)	0.000189(0.000034)	<.0001
NYHA C II($\hat{\alpha}_{25}$)	-0.2059 (0.03096)	<.0001	Dias($\hat{\alpha}_{35}$)	0.000196(0.000050)	<.0001
NYHA C III($\hat{\alpha}_{26}$)	-0.1804 (0.02571)	<.0001	Smok($\hat{\alpha}_{36}$)	0.07155(0.02424)	0.0032
Sys($\hat{\alpha}_{27}$)	0.000913(0.000121)	<.0001	Sigma1(σ_3)	0.000350(0.000040)	<.0001
Dias($\hat{\alpha}_{28}$)	0.000644(0.000177)	0.0003	ε_3	0.6728(0.03860)	<.0001
Vhds($\hat{\alpha}_{29}$)	0.03979(0.01444)	0.0059	$\hat{\sigma}^2_{b30}$	0.003277(0.000393)	<.0001
			Common parameters		
NYHA C*Ti I*T($\hat{\alpha}_{210}$)	0.07791(0.01139)	<.0001	$\sigma_{b20,b30}$	-0.00082(0.000646)	0.2060
NYHA C*Ti II*T($\hat{\alpha}_{211}$)	0.06973 (0.01142)	<.0001			
NYHA C*Ti III*T($\hat{\alpha}_{212}$)	0.05675(0.009663)	<.0001	$\sigma_{b30,b31}$	0.000258(0.000258)	0.3180
Sigma1(σ_2)	0.002456 (0.000245)	<.0001			
ε_2	0.4342 (0.05536)	<.0001			
$\hat{\sigma}^2_{b20}$	0.01237(0.002149)	<.0001			
$\sigma_{b20,b21}$	-0.00397(0.000802)	<.0001			
$\hat{\sigma}^2_{b21}$	0.001647(0.000339)	<.0001			

C)

39

4.4.3 Multivariate model

4.4.3.1 Selecting fixed effects for the multivariate model

To select significant variables backward variable selection method was used for also multivariate model as the above two models. The same procedure was used as them to select the fixed effects of the bivariate models.

Finally a model with fixed effects sqrtim1, sqrtim2, tim3, int1, int2, int3, sex1, sex2, sex3, age1, age2, age3, NYHAC1, NYHAC2, sys1, sys2, sys3, diast1, diast2, diast3, diag2, smoks3, vhds2, sqrtim1*NYHAC1 and sqrtim2*NYHAC2 was selected as a final multivariate model for pulse rate, Respiratory rate and weight with relatively small values of AIC= -12119.8, BIC= -12010.4 and log-Likelihood ratio test with P-value of <0.0001.

4.4.3.2 Selecting the covariance structure of random effects for the multivariate model

According to **table 14**; the final multivariate model for random effects with unstructured covariance structure was preferred with the relatively small value of AIC= **-12466.8**, BIC= **-12418.2** and -2LL= **-12498.8** of the model. To identify this best variance-covariance structure of random effects REML estimation method was used.

4.4.3.3 Multivariate model selection for correlation structure of errors

According to the **table 8;** the most use-full correlation structures of error is Autoregressive for the multivariate model since a model with this correlation structure has the smallest AIC values, BIC values, and the Log Likelihood scores compared to the other models. To identify this best variance-covariance structure of error REML estimation method was used with the selected fixed effects and random effects with unstructured variance-covariance structure model which was selected as a final best model.

Table 8: Correlation structures of measurement errors for multivariate model

Multivariate model	AIC	BIC	-2LL
Compound symmetry	-12464.8	-12413.2	-12498.8
Variance components	-13514.4	-13459.7	-13550.4
Autoregressive	**-13959.2**	**-13895.4**	**-14001.2**

4.4.3.4 The results of the final multivariate mixed effect model

A multivariate mixed effect model for the three symptoms of CHF Pr, Rr, and Wh were fitted with an unstructured covariance and Autoregressive correlation structure. This model is the same as the bivariate models by considering the association between the pulse rate, respiratory rate and weight but the difference in this model is considering the association of pulse rate, respiratory rate and weight at the same time which increases the number of the parameters. In a joint multivariate model of the log of pulse rate, a log of respiratory rate and a log of weight age, time, sex, systolic blood pressure, and diastolic blood pressure are common significant factors. The New York Heart Association Class and its interaction effect with time for both a log pulse rate and a log of respiratory rate, Valvular heart disease for log respiratory rate and Smoking for a log of weight are also significant factors.

The estimated parameters and its standard error of intercept for log of Pulse rate, log of Respiratory rate and log of weight are 1.9094(0.02248), 1.4438(0.03161) and 1.6785(0.01565) represents an average of log of Pulse rate, log Respiratory rate and log of weight during the first follow up time respectively and excluding all covariates in the model. From all covariates, age, systolic blood pressure and diastolic blood pressure were positively associated with the three outcomes that mean the increase of age, systolic and diastolic blood pressure leads to a particular increase on the three outcomes. The time had a negative effect (-0.05257(0.005444), -0.1175(0.007202) and -0.00209(0.000234)) on joint evolution of Pr, Rr, and Wh respectively. Sex had also significant effect on Pr, Rr, and Wh outcomes

Table 9: *Parameter estimates and standard errors for multivariate linear mixed effects model*

Multivariate	LogPr		LogRr			LogWh		
Effect	Estimate(s.e)	p-value	Effects	Estimate(s.e)	p-value	Effects	Estimate(s.e)	p-value
Inter($\hat{\alpha}_{10}$)	1.9094(0.02248)	<.0001	Inter($\hat{\alpha}_{20}$)	1.4438(0.03161)	<.0001	Inter($\hat{\alpha}_{30}$)	1.6785(0.01565)	<.0001
T($\hat{\alpha}_{11}$)	-0.05257(0.005444)	<.0001	T($\hat{\alpha}_{21}$)	-0.1175(0.007202)	<.0001	T($\hat{\alpha}_{31}$)	-0.00209(0.000234)	<.0001
Age($\hat{\alpha}_{13}$)	0.000998(0.000221)	<.0001	Age($\hat{\alpha}_{23}$)	0.000982(0.000335)	0.0033	Age($\hat{\alpha}_{31}$)	0.001515(0.000283)	<.0001
Sex($\hat{\alpha}_{14}$)	-0.02216(0.007066)	0.0017	Sex($\hat{\alpha}_{24}$)	-0.03449(0.01056)	0.0011	Sex($\hat{\alpha}_{32}$)	0.03430(0.009868)	0.0005
NYHAC I($\hat{\alpha}_{15}$)	-0.1245(0.02244)	<.0001	NYHAC I($\hat{\alpha}_{25}$)	-0.2099(0.03350)	<.0001	Sys($\hat{\alpha}_{38}$)	0.000189(0.000034)	<.0001
NYHAC II($\hat{\alpha}_{16}$)	-0.1430(0.02167)	<.0001	NYHAC II($\hat{\alpha}_{26}$)	-0.2008(0.03134)	<.0001	Dias($\hat{\alpha}_{37}$)	0.000196(0.000050)	<.0001
NYHAC III($\hat{\alpha}_{17}$)	-0.08198(0.01885)	<.0001	NYHAC III($\hat{\alpha}_{27}$)	-0.1821(0.02606)	<.0001	Smok($\hat{\alpha}_{38}$)	0.07749(0.02445)	0.0015
Sys($\hat{\alpha}_{18}$)	0.000984(0.000110)	<.0001	Sys($\hat{\alpha}_{28}$)	0.000906(0.000120)	<.0001	Sigma1(σ_3)	0.000352(0.000040)	<.0001
Dias($\hat{\alpha}_{19}$)	0.000693(0.000163)	<.0001	Dias($\hat{\alpha}_{29}$)	0.000659(0.000175)	0.0002	ε_3	0.6743(0.03869)	<.0001
NYHAC*Time I*T($\hat{\alpha}_{110}$)	0.03897(0.008600)	<.0001	Vhds($\hat{\alpha}_{210}$)	0.03390(0.01280)		$\hat{\sigma}^2_{b30}$	0.003368(0.000408)	<.0001
II*T($\hat{\alpha}_{111}$)	0.03545(0.008736)	<.0001				Common parameters		
III*T($\hat{\alpha}_{112}$)	0.02255(0.007431)	0.0024	NYHAC*Time I*T($\hat{\alpha}_{211}$)	0.08043(0.01153)	<.0001	Effects	Estimate(s.e)	P_value
Sigma1(σ_1)	0.001901(0.000132)	<.0001	II*T($\hat{\alpha}_{212}$)	0.06630(0.01173)	<.0001	$\sigma_{b10,b20}$	0.006873(0.001260)	<.0001
ε_1	0.3135(0.04615)	<.0001	III*T($\hat{\alpha}_{213}$)	0.05916(0.009703)	<.0001	$\sigma_{b10,b30}$	-0.00039(0.000481)	0.4123
$\hat{\sigma}^2_{b10}$	0.005817(0.001153)	<.0001	Sigma1(σ_2)	0.002288(0.000194)	<.0001	$\sigma_{b10,b21}$	-0.00263(0.000500)	<.0001
$\sigma_{b10,b11}$	-0.00201(0.000450)	<.0001	ε_2	0.3956(0.05097)	<.0001	$\sigma_{b20,b30}$	-0.00092(0.000679)	0.1765
$\hat{\sigma}^2_{b11}$	0.000889(0.000195)	<.0001	$\hat{\sigma}^2_{b20}$	0.01402(0.002260)	<.0001	$\sigma_{b20,b11}$	-0.00237(0.000515)	<.0001
			$\sigma_{b20,b21}$	-0.00464(0.000851)	<.0001	$\sigma_{b30,b11}$	0.000215(0.000205)	0.2952
			$\hat{\sigma}^2_{b21}$	0.001984(0.000361)	<.0001	$\sigma_{b30,b21}$	0.000336(0.000274)	0.2195
						$\sigma_{b11,b21}$	0.001120(0.000212)	<.0001

Where, inter=intercept, T=Time, Sys=systolic blood pressure, Dias= diastolic blood pressure, Vhds=valvular heart disease, and smok=smoking status

4.4.4 Comparison of separate, joint bivariate and multivariate mixed effect models

The separate models have fitted for the three outcomes together anyway, but assuming that $\rho = 0$ (fit as a joint model with appropriate covariance terms equal to zero), which is entirely equivalent to fitting the models separately. The age, time, sex, New York Heart Association Class, systolic blood pressure, diastolic blood pressure, interaction effects of New York Heart Association Class with time, Valvular heart disease, Smoking and the interaction effects of Diagnostic history with Time are significant factors for pulse rate, respiratory rate, and weight of congestive heart failure patients.

In a joint multivariate model of the log of pulse rate, a log of respiratory rate and a log of weight age, time, sex, systolic blood pressure and diastolic blood pressure are common significant factors. The New York Heart Association Class and its interaction effect with time for both pulse rate and respiratory rate, Valvular heart disease for respiratory rate and Smoking for weight are also significant factors. Based on the information criteria the multivariate model better than bivariate models, bivariate models are better than separate models. In general, the models provide approximately similar results for the fixed effect parameter estimates but their associated standard error for the joint models decreased than separate and almost similar for joint bivariate and multivariate models.

4.5 Model diagnosis

In this study, the transformed response variable was used to meet the assumption of the normality. Again the different diagnostic checking plots for the final separate linear mixed models of transformed response variables log of Pr, a log of Rr, and log of Wh are presented in **Figures 8, 9** and **10**. According to the Figures **8.A, 9.A** and **10.A**, the plot of residuals versus fitted, even if there are some outliers, it was indicated that the variability of the errors in the log of Pr, log of Rr and log of Wh are almost nearly constant. The residuals are symmetric around zero (i.e. positive and negative residuals are almost equal). The Figures **8.B, 9.B** and **10.B** of the outcomes log of Pr, log of Rr and log of Wh, respectively seems bell-shaped plot which supports the normality assumption of errors. The errors are normally distributed with mean zero and constant variance.

In the same way, based on the histograms for the random effects with subject-specific random intercepts and random slopes those are shown in Figures **8.C, 8.D, 9.C, 9.D,** and **10.C**, it seems slightly deviated from a bell-shape on the random slope (Time) for log of Rr and log of Wh that is not that much deviation. So, there is no more problem with normality assumptions of random intercepts and random slopes for a log of Pr, a log of Rr and log of Wh models and the normality assumption are almost satisfied.

4.6 Discussion on the result

The three models were considered for fitting three outcome variables of congestive heart failure patients which measured longitudinally. They were separate, bivariate and multivariate linear mixed effects models. The model building for more than one response variable usually starts from separate models for each component, initially, each data is analyzed separately. The separate analysis is preferred to specify the random and fixed effects to be included in the linear mixed effect model and it provides as a hint to the values to be obtained in the bivariate or multivariate models.

In a linear mixed model separate analysis of log of pulse rate, a log of respiratory rate and a log of weight were carried out. Before fitting the linear mixed model for each out-comes, exploring the data analysis is necessary and has been explored to understand the data structure and determine the relevant modeling approaches. The individual's profile plot indicates the existence of variability in all the three outcomes of the congestive heart failure within and between patients. The exploratory analysis result for mean structure (loess smooth curve) also suggested that on average, the measure of pulse rate, respiratory rate the same as *Fissuh and Muleta* [7] and weight had a decreasing evolution over time, but the rate of evolution in weight was lower than that of pulse rate and respiratory rate.

Most of the time the health-related data may have the problem to fail the assumptions of the models, the same in this data the responses were not normally distributed and linearly related with time this contradict the normality assumption of pulse rate and respiratory rate of the CHF patients and linearity assumption that; the time is linearly related with pulse rate and respiratory rate of the CHF patients done by *Fissuh and Muleta* [7]. To solve the problem of non-normality the log function was considered and for non-linearity the time was transformed

44

to the square root of time. The assumption of linearity was not failed to the log of the weight of the congestive heart failure patients, so time was not transformed to the square root of time.

The fixed and random effect components were selected to include in the models. After the selection of the appropriately fixed effects by using the maximum likelihood estimation method, a linear mixed model without random effect, intercept, intercept and sqrttime and intercept, sqrttime and time were fitted. The fitted random effect models were compared for the purpose of selecting the best random effects that enable to account the variability between congestive heart failure patients as *Negash et al.* [54] compared in the study of joint modeling of longitudinal systolic and diastolic blood pressure measurements of hyper-tensive patients receiving treatment and *Fissuh and Muleta* [7] compared in A Joint Model for a Longitudinal Pulse Rate and Respiratory Rate of Congestive Heart Failure Patients. The four models were compared using the AIC, BIC and -2LL value and we got a model with intercept and linear sqrttime effect as random effect is the best after transformation of time for log of pulse rate and log of respiratory rate to square root of time which contradict by the study of *Fissuh and Muleta* [7] in which time was linear without transformation.

The covariance structure of random effects which used in this study were; unstructured, compound symmetric, heterogeneous compound symmetric, Toeplitz, variance components and autoregressive covariance structure of order one and compared using the information criteria AIC, BIC, and -2LL. Based on the information criteria the unstructured covariance structure was the most appropriate covariance structure of the models for the random effects of separate, bivariate and multivariate models which is consistent by the study of *Fissuh and Muleta* [7] and *Thiébaut et al.* [38].

In order to model dependence among observations, the correlation structure of the measurement error was considered. The correlation structures which used in this study were compound symmetric; variance components   and autoregressive correlation structure of order one and compared to select the most appropriate one by using AIC, BIC and -2LL. The autoregressive structure of order one was the most appropriate correlation structure of the measurement error as (*Chi and Reinsel,* [55]; *Lindstrom and Bates*, [56]) incorporated in-to

mixed models to consider the dependence among the observations which contradict the independence assumptions of the measurement error in *Fissuh and Muleta* [7].

The final separate, bivariate and multivariate model was fitted with the same procedure as explained in the above. The information criterion for the multivariate model is less than the bivariate model is less than a separate model. Based on their information criteria the multivariate model is the best model than others as *Thiebauta et al.* [38], and *Bo and Sheng* [41]. There are differences in the average longitudinal evolutions with in age, sex, time, NYHAC, Systolic blood pressure, diastolic blood pressure smoking, valvular heart disease, and New York Heart Association Classes *time. The data were analyzed by SAS software and the missing data in this study was assumed MAR the same as *Fissuh and Muleta* [7].

5. Conclusion and recommendation

5.1. Conclusion

Three models considered in this study were separate, bivariate and multivariate for fitting three response variables measured longitudinally.

The common significant factors of pulse rate, respiratory rate and weight in separated models are age, time, sex, New York Heart Association Class, systolic blood pressure and diastolic blood pressure. The interaction effects of New York Heart Association Class with time are significant factors for both pulse rate and respiratory rate. The Valvular heart disease for respiratory rate and Smoking and the interaction effects of Diagnostic history with Time are significant factors for the weight of congestive heart failure patients in separated models.

In a joint model of the log of pulse rate, a log of respiratory rate and a log of weight age, time, sex, systolic blood pressure, and diastolic blood pressure are common significant factors. The New York Heart Association Class and its interaction effect with time for both pulse rate and respiratory rate, Valvular heart disease for respiratory rate and Smoking for weight are also significant factors in the joint model.

The multivariate model is the best model compared to the separate and joint bivariate models because its standard error of the parameter estimates is smaller. And also, the multivariate model has a very small AIC value which indicates that it fits the data better than the separate and joint bivariate models. In addition to random effects considering the autoregressive correlation structure for repeated effects needed for congestive heart failure data.

The evolution of Pr, Rr, and Wh decrease in a linear pattern over time after patients started the CHF treatments. There is evidence of differences in the average longitudinal evolutions with-in age, sex, time, NYHAC, Systolic blood pressure, diastolic blood pressure, smoking, valvular heart disease, and the interaction effect of New York Heart Association Classes with time.

The missing values issue is common in clustered or longitudinal data sets, especially in longitudinal studies due to dropout. The likelihood-based approach of PROC MIXED is that it can accommodate data that are missing at random.

5.2. Recommendation

The more focus on different health sector is providing different types of drugs for congestive heart failure patients. In nature for this type of disease-treating with only drug is not enough for patients under a follow-up clinic; also it is important to know factors that contribute to the progression of the pulse rate, respiratory rate and weight of the patients and providing the more counseling service how to control the effects of this factor.

The progression was found to be different in all patients due to age, sex, time, NYHAC, Systolic blood pressure, diastolic blood pressure smoking, and valvular heart disease. Further studies are required in the area of congestive heart failure with the necessary variables to identify the important risk factors and to improve the progression. Thus, the fitting multivariate model is recommended.

In many studies, no flexible correlation structures of measurement error were considered in modeling joint mixed models, but in some cases, it is necessary to consider correlation structures of measurement error in models because using independent correlation structures of measurement error model may display incorrect and inefficient inferences. In this study, it is focused on only three response variables with correlation structure of measurement errors which were compound symmetry, variance components and autoregressive order one, for future work, one might want to look at modeling more than three response variables over time with additional correlation structure of the measurement error.

6. References

1. Nadar S, Prasad N, Taylor RS, Lip GY., (2005),Positive pressure ventilation in the management of acute and chronic cardiac failure: a systematic review and meta-analysis. Int J Cardiol; 99(2):171–185.

2. Schober KE, Hart TM, Stern JA, et al., (2011) Effects of treatment on respiratory rate, serum natriuretic peptide concentration, and Doppler echocardiographic indices of left ventricular filling pressure in dogs with congestive heart failure secondary to degenerative mitral valve disease and dilated cardiomyopathy. *J Am Vet Med Assoc*.; 239(4):468–479.

3. American Heart Association (2002), Heart Disease and Stroke Statistics. Am Heart Assoc, Dallas.

4. Gorgas DL., (2014), Vital signs and patient monitoring techniques. In: Roberts JR, Hedges JR (eds.). Clinical Procedures in Emergency Medicine Sauders, Philadelphia, USA.

5. American Heart Association, (Feb 13, 2018), Warning Signs of Heart Failure. Updated.

6. Jennifer S. Wright; (September 30, 2017) Signs & Symptoms of Congestive Heart Failure in Women. Updated.

7. Fissuh YH, Muleta G (2015) A Joint Model for a Longitudinal Pulse Rate and Respiratory Rate of Congestive Heart Failure Patients: at Ayder Referral Hospital of Mekelle University, Tigray, Ethiopia. J Biom Biostat 6: 260. doi:10.4172/2155-6180.1000260

8. Sammel M, Lin X, Ryan L (1999) Multivariate linear mixed models for multiple outcomes. Stat Med 18:2479–2492

9. Pinheiro J, Liu C, Wu Y (2001) efficient algorithms for robust estimation in linear mixed-effects models using the multivariate t distribution. J Comput Graph Stat 10(2):249–276

10. Fieuws S, Verbeke G (2004) Joint modeling of multivariate longitudinal profiles: pitfalls of the random effect approach. Stat Med 23:3093–3104

11. Bandyopadhyay S, Ganguli B, Chatterjee A (2011), Review of the multivariate longitudinal data analysis. Stat Methods Med Res 20(4):299–330 E.E. Tripoliti et al. / Computational and Structural Biotechnology Journal 15 (2017) 26–47

12. Cook C, Cole G, Asaria P, Jabbour R, Francis DP., (2014), The annual global economic burden of heart failure. Int J Cardiol.; 171(3):368–76. doi:10.1016/j.ijcard.2013.12.028.

13. GBD 2015 Disease and Injury Incidence and Prevalence, Collaborators. (8 October 2016"). Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015"(*https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5055577*). *Lancet*.**388**(10053):1545–1602.doi:10.1016/S0140-6736(16)31678-6*(https://doi.org/10.1016%2FS0140-6736%2816%2931678-6).PMC5055577* (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5055577).PMID27733282 (*https://www.ncbi.nlm.nih.gov/pubmed/27733282*)

14. Metra, M; Teerlink, JR (28 October 2017). "Heart failure". *Lancet*. **390** (10106): 1981–1995.doi:10.1016/S0140-6736(17)310711 (https://doi.org/10.1016%2FS01406736%2817%2931071-1.)PMID28460827 (https://www.ncbi.nlm.nih.gov/pubmed/28460827)

15. McMurray JJ, Pfeffer MA (2005). "Heart failure".*Lancet*. **365** (9474): 1877–89. doi:10.1016/S0140-6736(05)66621-4 (https://doi.org/10.1016%2FS0140-6736%2805%2966621-4.)PMID15924986 (https://www.ncbi.nlm.nih.gov/pubmed/15924986).

16. Dickstein K, Cohen-Solal A, Filippatos G, et al. (October 2008)".E SC Guidelines for the diagnosis and treatment of acute and chronic heart failure 2008: the Task Force for the Diagnosis and Treatment of Acute and Chronic Heart Failure 2008 of the European Society of Cardiology Developed in collaboration with the Heart Failure Association of the ESC (HFA) and endorsed by the European Society of Intensive Care Medicine (ESICM)"(http://eurheartj.oxfordjournals.org/cgi/pmidlookup?view=long&pmid=18799522.)*Eur.HeartJ*.**29**(19):2388442.doi:10.1093/eurheartj/ehn309(https://doi.org/10.1093%2Feurheartj%2Fehn309.)PMID18799522(https://www.ncbi.nlm.nih.gov/pubmed/18799522)

17. Mayo Foundation (Dec, 23, 2017), Medical Education and Research. https://www.mayoclinic.org/diseases-conditions/heart-failure/symptoms-cause/syc-20373142.

18. Michael S Figueroa MD and Jay I Peters MD FAARC 2006, Congestive Heart Failure: Diagnosis, Pathophysiology, Therapy, and Implications for Respiratory Care

19. Belete Fessahaye Alemseged. (2010) The pattern of cardiac diseases at the cardiac clinic of Jimma university specialized Hospital, South West Ethiopia. Ethiop J Health Sci, Vol 20, Pp.99-105.

20. American Heart Association, (Feb 13, 2018), Warning Signs of Heart Failure, Updated.

21. American Heart Association, (2015), Causes of heart failure.

22. Molenberghs, G. and Verbeke, G. (2008). Review on linear mixed models for longitudinal data, possibly subject to drop out. *Statistical Modeling.* **1**(4), 235-269.

23. McCulloch, C., Searle, S. and Neuhaus, J. (2008). Generalized, Linear, and Mixed Models: John Wiley & Sons, Inc.

24. Nonhlanhla, Y. (2009). Modeling CD4+ Count Over Time in HIV Positive Patients Initiated on HAART in South Africa Using Linear Mixed Models. MSC

25. Pinheiro JC and Bates DM (2000). Mixed-Effects Models in S and S-PLUS. Springer Verlag New York.

26. Browne, W. & Goldstein, H. (2010). MCMC sampling for a multilevel model with non-independent residuals within and between cluster units. *Journal of Educational and Behavioral Statistics, 35*(4), 453-473. doi:10.3102/1076998609359788

27. Lloyd J. Edwards, PhD*, (2000), Pediatric Pulmonology 30:330–344.

28. Marie Davidian, (Feb, 2014), Stat Methods Med Res. 23(1): 42–59

29. Xu J, Zeger SL (2001), Evaluation of multiple surrogate endpoints. Biometrics 57:81–87

30. He B, Luo S (2013) Joint modeling of multivariate longitudinal measurements and survival data with applications to Parkinson's disease. Stat Methods Med Res (In Press Published on line first on April 16, 2013)

31. Adams R, Wilson M, Wu M (1997). Multilevel item response models: an approach to errors in variables regression. J Educ Behav Stat 22(1):47–76

32. Zwinderman A. (1991), Generalized Rasch model for manifest predictors. Psychometrika 56(4):589–600

33. Maier K (2001) A Rasch hierarchical measurement model. J Educ Behav Stat 26(3):307–330

34. Fox J (2005) Multilevel IRT using dichotomous and polytomous response data. Br J Math Stat Psychol 58(1):145–172

35. Catalano PJ, Ryan LM (1992). Bivariate latent variable models for clustered discrete and continuous outcomes. J Am Stat Assoc 87: 651-658.

36. Fitzmaurice G, Laird N (1995), Regression models for a bivariate discrete and continuous outcome with clustering'. J AM Stat Assoc 90: 845-852.

37. Thiébaut R, Jacqmin-Gadda H, Chêne G, Leport C, Commenges D (2007) Bivariate Linear Mixed Models Using SAS Proc MIXED.

38. Ferrari SLP, Cribari-Neto F (2004) Beta regression for modeling rates and proportions. J App Stat 31: 799-815.

39. Njagi EN, Molenberghs G, Verbeke G, Aerts M (2013), Joint Models for Survival and Longitudinal Data, Missing Data, and Sensitivity Analysis, with Applications in Medical Research.

40. Bo He and Sheng Luo, 2016, joint modeling of multivariate longitudinal measurements and survival data with application to Parkinson's disease

41. Lambert P, Vandenhende F (2002) A copula-based model for multivariate non normal longitudinal data: analysis of a dose titration safety study on a new antidepressant. Stat Med 15: 3197-3217.

42. Forough Pazhuheian and Farid Zayeri, (winter, 2018) Effect of training after discharge on readmission and re-hospitalization of patients with heart failure (randomized single-blind clinical trial), Journal of Paramedical Sciences (JPS), Vol 9, No1. ISSN 2008-4978

43. Badiru AB. (1992) A computational survey of univariate and multivariate learning curve models. Engineering Management, IEEE Transactions on. 39(2):176-88.

44. Santos AA, Nogales FJ, Ruiz E. (2013), Comparing a univariate and multivariate models to forecast portfolio value-at-risk. Journal of financial econometrics. 11(2):400-41.

45. Abramowitz M, Stegun IA (1964) Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. Dover, New York.

46. Pinheiro, J., and Bates, D. (2000). Mixed-Effects Models in S and S-PLUS. Springer Verlag New York.

47. Olkin, I. and Tate, R. (1961). Multivariate correlation models with mixed discrete and continuous variables. *Ann of Math Stat.* **32**, 448-465

48. Box, G., Jenkins, G, and Reinsel, G. (1994). Time Series Analysis: Forecasting and Control. Holden-Day

49. Diggle, P., Zeger, S., and Liang, K., (1994). Analysis of Longitudinal Data: University Press Inc., New York.

50. Shah, A., Laird, N. and Schoenfeld, D., (1997). A random-effects model for multiple characteristics with possibly missing data.**92**:775–779.

51. Sakamoto, Y., Ishiguro, M. and Kitagawa, G. (1986). Akaike Information Criterion Statistics. D. Reidel Publishing Company

52. Laird, N., and Ware, J. (1982). Random-effects models for longitudinal data. Biometrics

53. Negash et al.,(2016),Joint modeling of longitudinal systolic and diastolic blood pressure measurements of hyper-tensive patients receiving treatment

54. Chi, E. M., and Reinsel, G. C. (1989). Models for longitudinal data with random effects and AR (1) errors, *Journal of the American Statistical Association* **84**: 452–459.

55. Lindstrom, M. J., and Bates, D. M. (1990). Nonlinear mixed effects models for repeated measures data, *Biometrics* **46**: 673–687.

APPENDEX I

Table 10: Patterns of sample size

Time	1	2	3	4	5	6	7	8	9	10
size	154	154	154	151	141	122	106	86	75	55
Time	11	12	13	14	15	16	17	18	19	
size	42	33	24	15	11	8	6	5	4	

Table 11: Correlation structure of measurement errors for the separated models

	Models	AIC	BIC	-2LL
For pulse rate	Compound symmetry	-3808.0	-3792.8	-3818.0
	Variance components	-3810.0	-3797.9	-3818.0
	Autoregressive	**-3865.6**	**-3850.4**	**-3875.6**
For respiratory rate	Compound symmetry	-3502.1	-3486.9	-3512.1
	Variance components	-3504.1	-3492.0	-3512.1
	Autoregressive	**-3578.9**	**-3563.7**	**-3588.9**
For weight	Compound symmetry	-6014.5	-6005.4	-6020.5
	Variance components	-6016.5	-6010.5	-6020.5
	Autoregressive	**-6311.2**	**-6302.1**	**-6317.2**

Table 12: Variance-covariance structures of random effects for bivariate models

	Model	AIC	BIC	-2LL
pulse rate and respiratory rate	Compound symmetry	-6873.6	-6864.5	-6879.6
	Heterogeneous compound symmetry	-7095.3	-7077.1	-7085.5
	Variance components	-7075.9	-7060.7	-7085.9
	Unstructured	**-7403.4**	**-7370.0**	**-7425.4**
	Autoregressive	-6897.5	-6888.4	-6903.5
	Toeplitz	-6964.6	-6949.4	-6974.6
respiratory rate and weight	Compound symmetry	-8558.0	-8482.1	-8608.0
	Heterogeneous compound symmetry	-8693.1	-8611.1	-8747.1
	Variance components	-8639.8	-8560.9	-8691.8
	Unstructured	**-8816.6**	**-8728.5**	**-8874.6**
	Autoregressive	-8522.4	-8446.5	-8572.4
	Toeplitz	-8579.4	-8500.5	-8631.4
pulse rate and weight	Compound symmetry	-8747.1	-8737.9	-8753.1
	Heterogeneous compound symmetry	-8852.3	-8837.1	-8862.3
	Variance components	-8811.8	-8799.7	-8819.8
	Unstructured	**-8977.4**	**-8956.2**	**-8991.4**
	Autoregressive	-8722.6	-8713.5	-8728.6
	Toeplitz	-8784.8	-8772.7	-8792.8

Table 13: *Correlation structures of measurement errors for bivariate models*

Model			AIC	BIC	-2LL
pulse rate and respiratory rate		Compound symmetry	-7401.4	-7365.0	-7425.4
		Variance components	-7403.9	-7367.4	-7427.9
		Autoregressive	**-7523.9**	**-7481.3**	**-7551.9**
respiratory rate and		Compound symmetry	-8814.6	-8723.5	-8874.6
		Variance components	-9774.4	-9683.3	-9834.4
		Autoregressive	**-10178.7**	**-10081.6**	**-10242.7**
pulse rate and weight		Compound symmetry	-8989.0	-8964.7	-9005.0
		Variance components	-9884.3	-9860.0	-9900.3
		Autoregressive	**-10242.6**	**-10212.3**	**-10262.6**

Table 14: *Variance-covariance structures of random effects for the multivariate model*

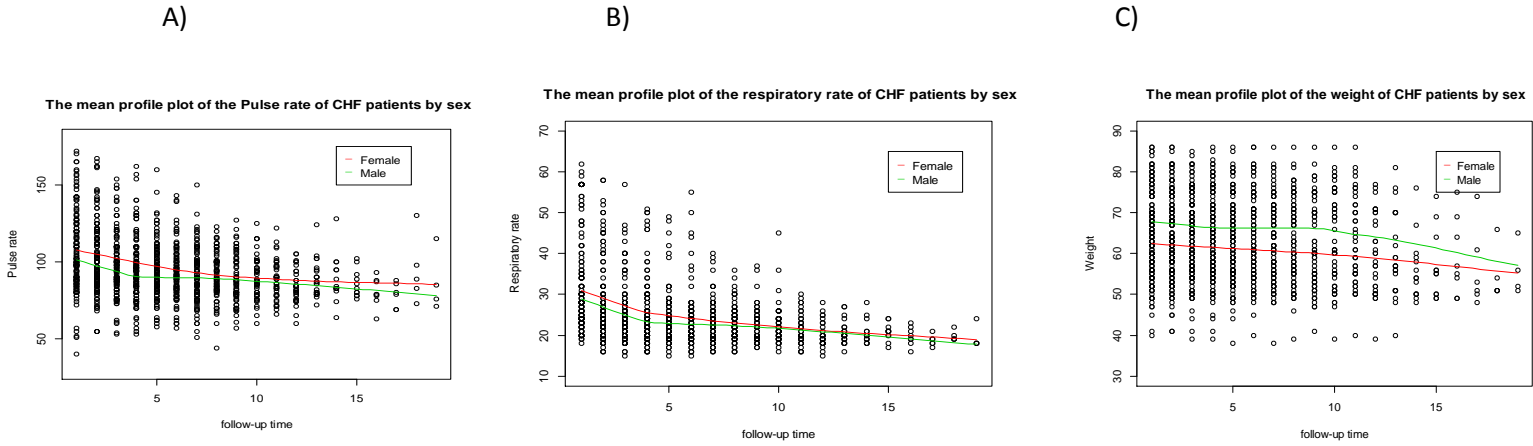| Model	AIC	BIC	-2LL
For Pr, Rr, and Wh			
Compound symmetry	-11838.7	-11829.6	-11844.7
Heterogeneous compound symmetry	-12083.3	-12062.0	-12097.3
Unstructured	**-12466.8**	**-12418.2**	**-12498.8**
Autoregressive	-11857.2	-11848.1	-11863.2
Toeplitz	-11969.1	-11950.8	-11981.1
Variance components	-12053.2	-12035.0	-12065.2

APENDEX II

A)
B)
C)

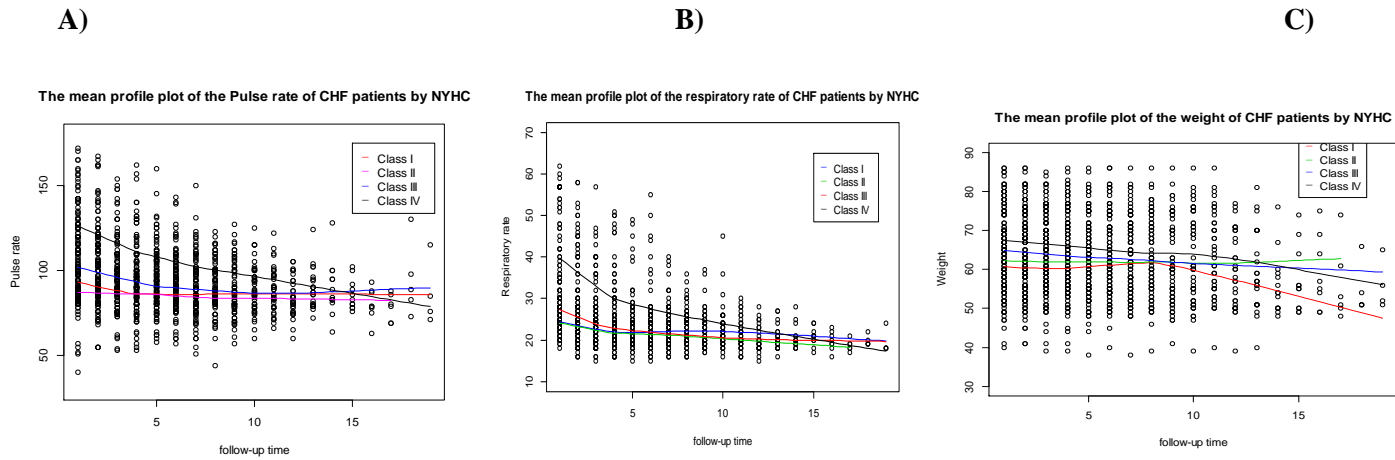Figure 3: Mean profile plot for Pr, Rr, and Wh of CHF patients by sex

A)
B)
C)

Figure 4: Mean profile plot for Pr, Rr, and Wh of CHF patients by NYHA
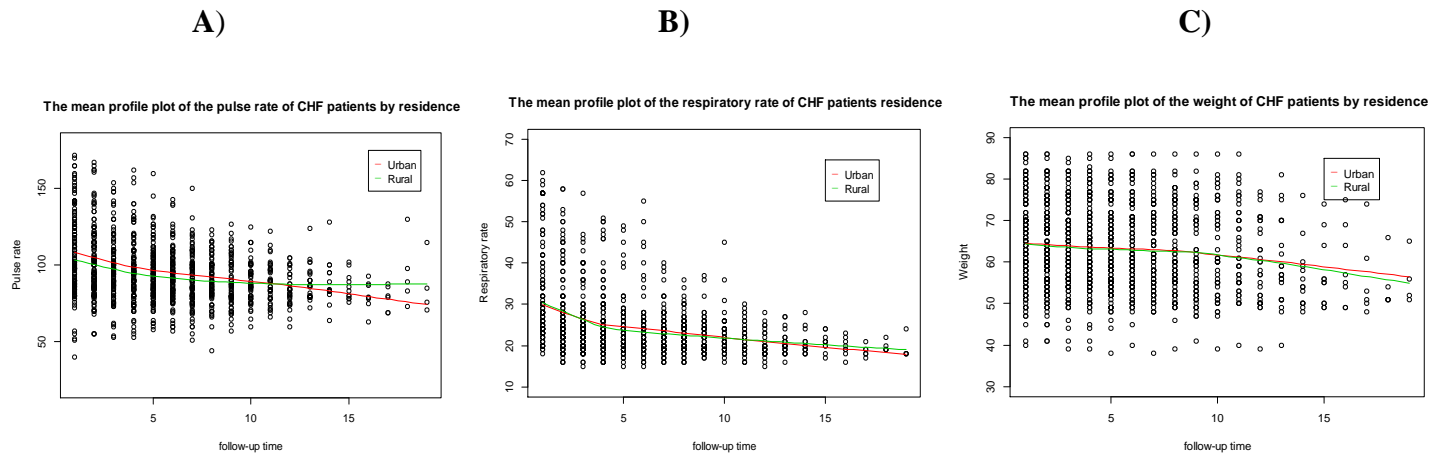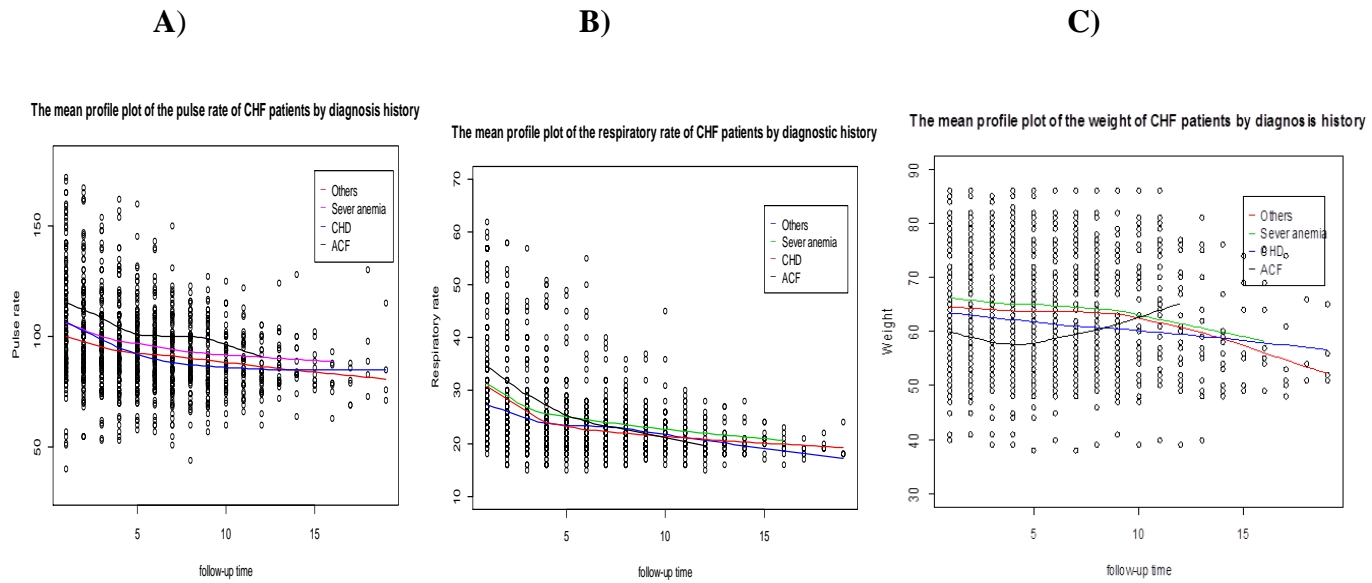
The mean profile plot of the pulse rate of CHF patients by residence

The mean profile plot of the respiratory rate of CHF patients residence

The mean profile plot of the weight of CHF patients by residence

Figure 5: Mean profile plot for Pr, Rr, and Wh of CHF patients by residence

The mean profile plot of the pulse rate of CHF patients by diagnosis history

The mean profile plot of the respiratory rate of CHF patients by diagnostic history

The mean profile plot of the weight of CHF patients by diagnosis history

Figure 6: Mean profile plot for Pr, Rr, and Wh of CHF patients by diagnostic history

A)   B)   C)

Variance structure of Pr   **Variance structure of Rr**   **Variance structure of Wh**

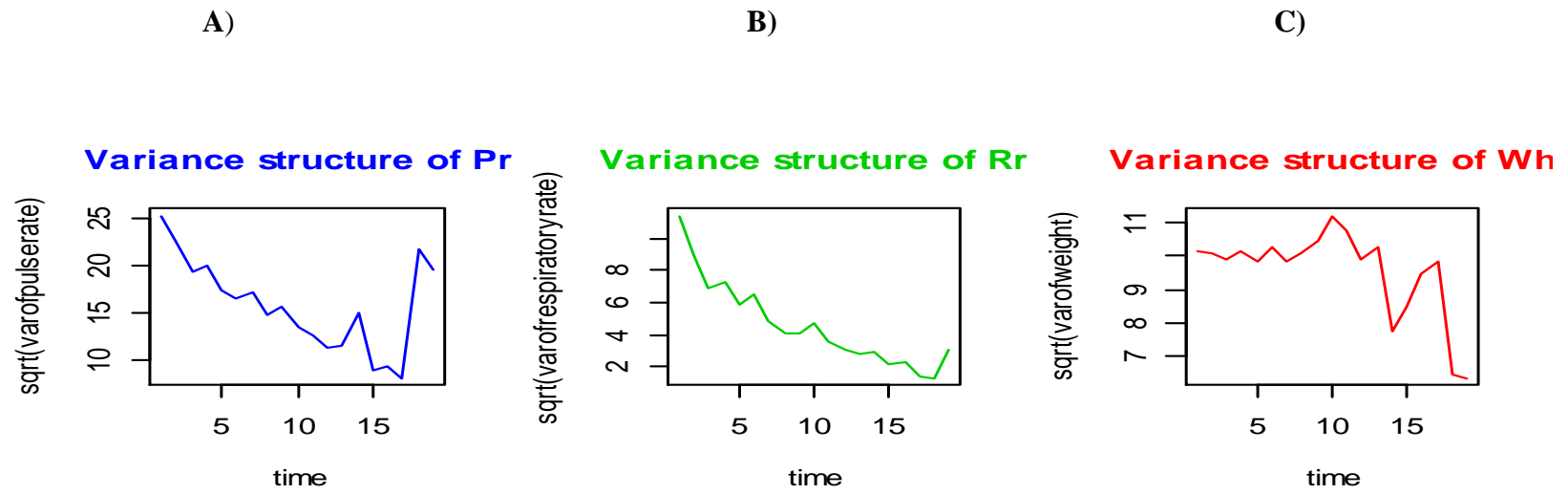Figure 7: Variances structure for Pr, Rr, and Wh of CHF patients
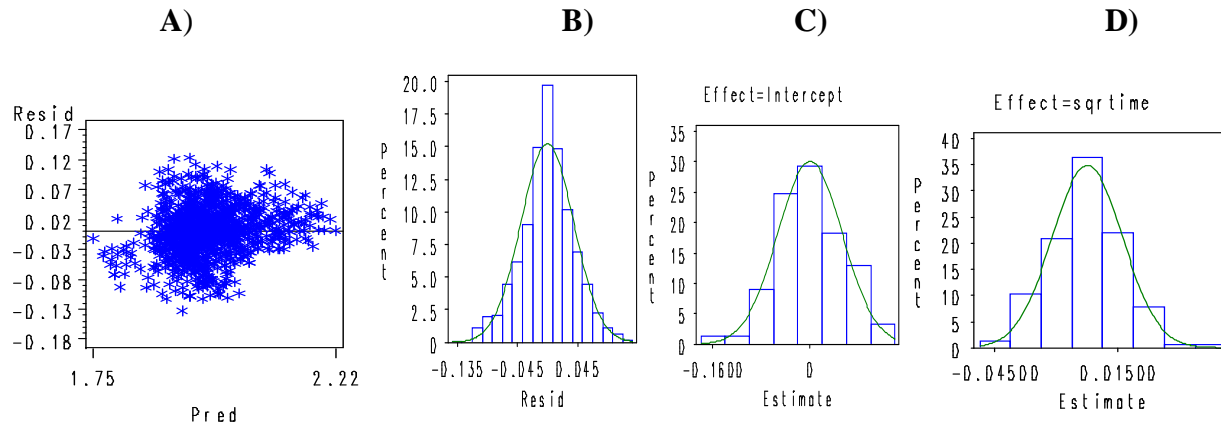

A)   B)   C)   D)

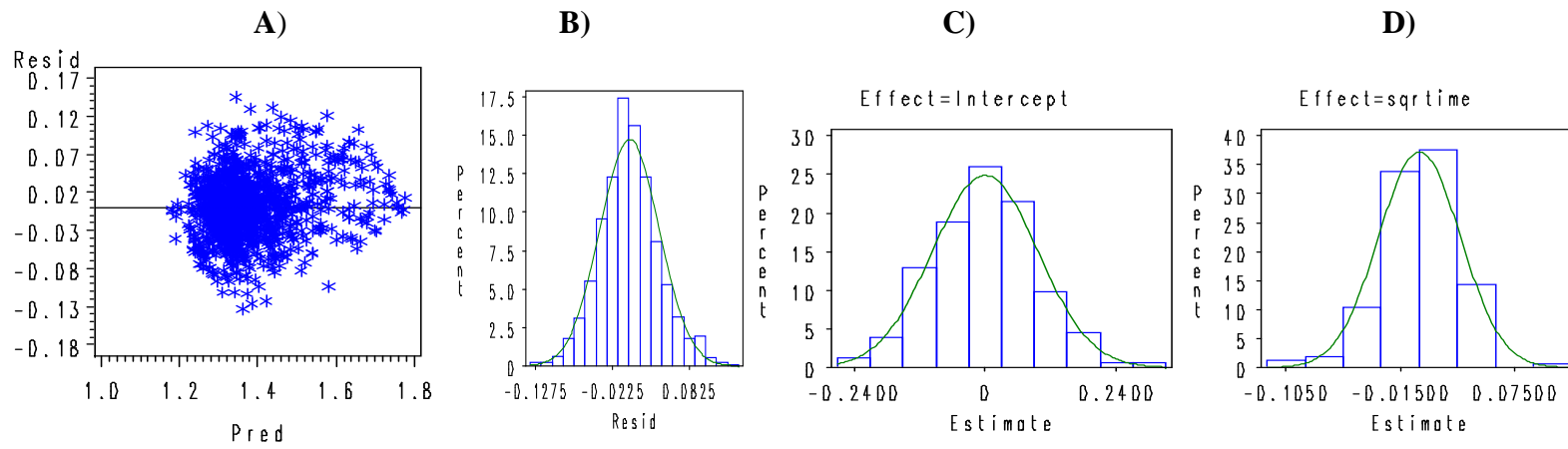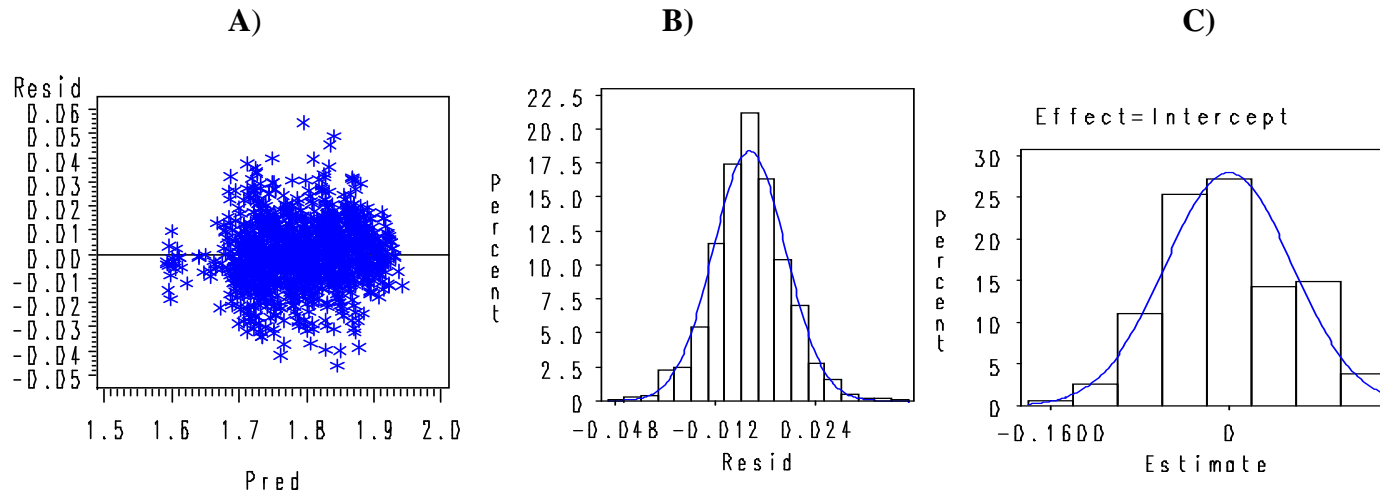Figure 8: Model checking for a log of pulse rate

Figure 9: Model checking for a log of respiratory rate

Figure 10: Model checking for a log of weight