CHAPTER 3: EMPIRICAL INVESTIGATION

In this chapter the general and specific research objectives of the empirical investigation is discussed. The discussion on the research methodology follows. This includes discussions on pharmacoepidemiology, drug utilization review (DUR) and cross-sectional studies. Topics of discussion that follow are the database, study population, classification systems, measurements, data analysis and statistical analysis. The reliability and validity of the research instruments are discussed, as well as the ethical considerations, discussion of the results of the empirical investigation, conclusion and recommendations and finally the chapter summary is given.

3.1 AIM

The aim of this study is to identify, interpret and analyse the prescribing patterns of warfarin and to evaluate the prevalence of the co-prescribing of drugs that has a potential interaction with warfarin.

3.2 RESEARCH OBJECTIVES

The research objectives of this study are divided into a general objective and specific objectives.

3.2.1 GENERAL RESEARCH OBJECTIVE

The general objective of this study is to compare the prescribing patterns, as well as the prevalence of potential drug-drug interactions on warfarin prescriptions in a South African private health care setting according to a pharmaceutical benefit management company situated in South Africa. Data were extracted for a timeframe of six years from 1 January 2005 to 31 December 2010.

3.2.2 SPECIFIC OBJECTIVES

The research project consisted of two phases, namely a literature review and an empirical investigation. The literature review was discussed in Chapter 1 (refer to section 1.4.1). The specific research objectives include the following:

3.2.2.1 Empirical Investigation

The specific research objectives of the empirical investigation include the following:

➤ To investigate the prescribing patterns of warfarin in a part of the private health care sector of South Africa;

➤ To investigate the prevalence of co-prescribing of potentially interacting drugs with warfarin in South Africa over a six year period, 2005-2010;

➤ To assess whether warfarin treatment is changed when prescribed together with potentially interacting drugs.

3.3 RESEARCH METHODOLOGY

This a cross-sectional, observational or quantitative study and the empirical investigation was done retrospectively on warfarin products.

3.3.1 RESEARCH DESIGN

The following concepts will be discussed in the research design: pharmacoepidemiology, drug utilization review (DUR), the necessity of DUR and the different types of DUR.

3.3.1.1 Pharmacoepidemiology

According to Strom (2006), the term pharmacoepidemiology is composed of two parts, namely "pharmaco" and "epidemiology". They believe that the "pharmaco" part of pharmacoepidemiology is related to clinical pharmacology, which is "the study of the effects of drugs in humans" (Strom, 2006:4, 5). The second part of pharmacoepidemiology is "epidemiology", which is "the study of the frequency, distribution and determinants of disease in humans" (Hertzema *et al.*, 2008:1). After understanding these two concepts, a formal definition can be given for pharmacoepidemiology. The WHO defines pharmacoepidemiology as "the study of the use and effects/side-effects of drugs in large numbers of people with the purpose of supporting the rational and cost-effective use of drugs in the population, thereby improving health outcomes" (WHO, 2003:8).

In order to simplify the main aim of pharmacoepidemiological research, Waning and Montagne (2001) identified three focus points for this category of research. They include an assessment of which drugs are being used for a specific condition, what the patterns of drug use are, and the reason for using these drugs (Waning & Montagne, 2001:5).

3.3.1.2 Drug utilisation review (DUR)

DUR stems from the idea of drug utilisation research, which the WHO defined in 1977 as "the marketing, distribution, prescription, and use of drugs in a society, with special emphasis on the resulting medical, social and economic consequences" (WHO, 2003:8).

3.3.1.3 The necessity of DUR

The primary goal of DUR is to eventually gather information that will enable rational drug use in a population (WHO, 2003:8, 9). It provides us with important information on the factors associated with drug use:

> ➤ The patterns of drug use – trends and costs that are associated with drug use;
> ➤ The quality of drug use – compares actual drug use to prescribing guidelines and drug formularies (WHO, 2003:8, 9);
> ➤ The determinants of drug use – includes all the factors that influence drug use such as traditional medical practices, patient expectations, organisation and economic state of health services, pricing of medicines, regulations on medicines and reimbursement systems (Laporte et al., 1993:513; WHO, 2003:8, 9);
> ➤ The outcomes of drug use – which include health outcomes and economic consequences (WHO, 2003:8, 9).

A better understanding of these factors means that guidelines can be set in place to enable rational drug use in a population (Fulda et al., 2004:440).

3.3.1.4 Types of DUR studies

DUR studies can be classified into two categories namely descriptive studies and analytical studies. Descriptive studies aim to describe the patterns of drug use and to identify problem areas where further study is necessary. Analytical studies aim to connect data on drug use to statistics on morbidity, treatment outcomes and quality of care to eventually evaluate if drug therapy is rational or not (WHO, 2003:8).

This study qualifies as a descriptive DUR study, as the patterns of warfarin use and prevalence of the co-prescribing of drugs with potential DDIs with warfarin ("problem areas") are described.

DUR studies can also be either quantitative or qualitative. Lee *et al.* (2006:400) states that quantitative studies aim to "quantify the present state, developmental trends, and time course of drug usage at various levels of the health care system, whether national, regional, local, or institutional." They further to say that qualitative studies are conducted to "assess the appropriateness of drug use, usually by linking prescription data to the reasons for the drug prescribing."

This study qualifies as a quantitative study, as warfarin prescribing patterns and use of warfarin with co-prescribed drugs are studied to see whether there are any trends in the prescription of these drugs.

DUR studies can be divided into three further categories. These categories are as follows:

- **Retrospective studies:** the events of interest have already transpired, for example the drugs have already been prescribed by the physician, dispensed by the pharmacist and used by the patient. The data collected by these studies are reviewed and analysed in order to collect drug use data in a specific environment (Kidder & Bae, 1999:114; Truter, 2008:95).

  As stated in the beginning of section 3.3, the empirical study on warfarin products was done retrospectively.

- **Concurrent studies:** these studies are conducted while the drug is being dispensed. The advantage of this type of study is that when a potential problem occurs, it can be dealt with before the drug is dispensed and in this manner any serious problems that might harm the patient can be prevented (Guo *et al.,* 1995:1175; Truter, 2008:95).

- **Prospective study:** ideally a prospective study should be conducted by the prescribing physician. After taking a thorough medical history and obtaining complete set of medical documents of the patient, the prescribing physician can then make an informed decision on the therapy of the patient. The physician can then know which drugs to prescribe and whether drug therapy should be discontinued (Fulda *et al.,* 2004:433; Truter, 2008:95).

3.3.1.5 Cross-sectional studies

Cross-sectional studies form part of a group of study designs also known as observational studies (Mann, 2003:54). Observational studies serve the purpose of just that, observation. The researcher or "investigator" only observes how certain patterns evolve depending on the type of study that is conducted; therefore no intervention is carried out (Mann, 2003:54).

Cross-sectional studies are conducted to assess the prevalence of a certain outcome for a given population. It is usually a descriptive study that describes a population according to a specific outcome (Levin, 2006:24). There is no hypothesis testing involved. Cross-sectional studies assess the prevalence for a specific outcome for a specific population in one point in time for a short period (Levin, 2006:24).

There are several advantages when conducting a cross-sectional study. The two main advantages of cross-sectional studies are that it is quite inexpensive and faster to conduct compared to other study designs. With cross-sectional studies the prevalence of disease, risk factors and exposures can be assessed (DiPietro, 2010:977). Kang *et al.* (2005) also states that several exposures or diseases can be studied at the same time.

There are several disadvantages when conducting cross-sectional studies. Temporal associations or events that occur at the same time but have no relation to one another cannot be established. Exposure and outcome data are collected simultaneously, thus it is not possible to assess whether the exposure followed the outcome. Cross-sectional studies do not serve the function of studying diseases that are rare or diseases that occur over a short period of time. Fatal cases may be excluded because the time of data collection could have occurred before these fatalities, thus survival bias may be introduced (DiPietro, 2010:977; Kang *et al.*, 2005:139).

3.3.2 DATABASE

The empirical investigation on warfarin products was done retrospectively on warfarin products. These warfarin products were claimed through a medicine claims database of a pharmaceutical benefit management company. The time period of data analyses from the medicine claims database is from 1 January 2005 to 31 December 2010. Drug products that were co-prescribed with warfarin were also analysed as data from the medicine claims database. Data was also analysed for the total number of prescriptions of all drug products in the study period. According to the database the total number of prescriptions that were claimed through the database was

49 523 818. The total number of medicine items on these prescriptions were 118 305 941 and the total number of patients on the database was 7 748 621. This set of data was used to compare the data of warfarin prescriptions. Statistical analyses were done on the data to obtain the results. This was done by means of the SAS 9.1. ® Computer package (SAS Institute, 2004).

The following steps were used in the performance of the empirical investigation:

- Selection of the data source and identification of the study population with reference to the presented data and data analyses;
- Selection of the research design and identification of the research instruments. This is very important as this is the guide to show what is being studied. The primary focus of this study was warfarin and the prevalence of co-prescribed drugs that have a potential interaction with warfarin, as discussed in chapter 4;
- Data analyses and management were done by means of statistical methods. Data analyses employed statistical concepts in order to analyse the data according to the measuring instruments chosen;
- Data collection was done directly from the medicine claims database and it is the belief of the researcher that these data are correct and precise and therefore reliable and valid;
- The findings and the results of the empirical investigation will be discussed in chapter 4;
- From the findings and results obtained in the empirical investigation, conclusions will be drawn, recommendations will be made and the limitations of the study will be discussed in chapter 5.

3.3.2.1 Fields

The following fields were used to extract the required data from the database.

- Member information:
  - ➢ There were no patient names disclosed in the database. Each member was given a random number from which the patient could be identified. The date of birth of the patient was also available to the researcher. The age of the patient on the day of treatment was calculated from the first day of the next year. Other information that was also available was the gender of the patient (refer to section 3.3.4).

- Prescriber information:

➤ As with the patients, each prescriber was allocated a random number from which the prescriber could be identified. Information of the specific practice of the prescriber was not available. Different categories of prescribers could be identified (refer to section 3.3.4).

- The following medicine information was available on the database:
  ➤ The trade name/active ingredient of warfarin products, as well as other medicine prescribed together with warfarin;
  ➤ The NAPPI (National Pharmaceutical Product Interface) is part of a unique national coding system. These "codes" are allocated to all consumable pharmaceutical, surgical and other healthcare products in South Africa. All medical schemes recognise these codes, thus all the products that contain these codes are covered by medical schemes. This coding system aids in the electronic transfer of information throughout the healthcare chain (Anon., 2012). NAPPI codes are 9-digit numbers that include information such as the product name, pack size, strength, manufacturer and other information regarding the medicine product (Snyman, 2011:7a);
  ➤ Amounts prescribed and days supplied to determine the prescribed daily dose (PDD) – refer to section 3.3.5.2;
  ➤ The strength of each drug was programmed in the SAS 9.1 ® computer package and therefore could also be extracted from the database.

3.3.3 STUDY POPULATION

The study population consisted of the total data collected from the medicine claims database from 1 January 2005 until 31 December 2010. From the total data the number of warfarin prescriptions was extracted.

3.3.4 CLASSIFICATION SYSTEMS

The following classification systems were used to classify medicine items claimed through the medicine claims database:

- **Age groups** – the following age groups were selected in accordance with the age groups that were examined in the literature (refer to section 2.2.5.3):
  ➤ Age group 1 = >0 to ≤ 20 years
  ➤ Age group 2 = >20 to ≤ 39 years
  ➤ Age group 3 = >39 to ≤ 59 years

> Age group 4 = >59 years

- **Gender** – the gender of the patients was classified into three categories namely female, male and unspecified.

- **Prescriber** – four types of prescribers were identified on the database namely general practitioners, cardiologists, internists and other. The fourth type of prescriber, classified as "other", includes all the prescribers that do not fit into the previous groups.

- **Drug-drug interactions (DDI)** – the potential DDIs of the drugs co-prescribed with warfarin was classified according to the significance of the drug interaction as described by Tatro (2011). According to Tatro (2011), drug interactions can be classified according to the severity and documentation level of the drug interaction (refer to section 2.3.3).

- **MIMS classification** - Warfarin is classified under anticoagulants, which are further classified under section 8.2 of blood and haemopoeitic agents according to the Monthly Index of Medical Specialities (MIMS) (Snyman, 2011:13a).

3.3.5 RESEARCH MEASUREMENTS

The following research measurements were used during the empirical investigation:

3.3.5.1 Prevalence

Prevalence is "the number of all new and old cases of a disease or occurrences of an event during a particular period" (Mosby, 2009:1512). Prevalence should not be confused with incidence, which is "the frequency (number) of new occurrences of disease, injury, or death" (Jekel et al., 2007:38). As stated in the aim of the study (refer to section 3.1), the study focuses on the prevalence of the co-prescribing of drugs that could potentially have DDI with warfarin.

3.3.5.2 Prescribed daily dose (PDD)

As far back as the 1970s researchers realised that a common unit of measurement was needed in order to perform drug utilisation studies in different countries (Mantel-Teeuwisse et al., 2001:1181; Truter et al., 1996:675). As different countries used different units of measurement, the drug utilisation studies from different countries could not be compared to one another, thus the defined daily dose (DDD) was created. By using the DDD as a unit of measurement in drug utilisation studies, studies conducted in different countries around the world could now be compared to one another (Cosentio et al., 2000:513; Truter et al., 1996:675: WHO, 2003:6, 7).

The WHO (2003:38) defines the DDD as the "assumed average maintenance dose per day for a drug used for its main indication in adults." The DDD of a drug is usually determined by examining medical literature, the information given in the drug manufacturers' data sheet and the experience of the use of the product (Truter et al., 1996:67; WHO, 2003:38).

The DDD can be expressed in three ways:

- DDDs per 1000 inhabitants per day – this illustrates the proportion of the study population that is treated daily with a particular drug or group of drugs (Capellá, 1993:57).
- DDDs per 100 bed-days – this is used in a hospital setting where drugs are used by inpatients (WHO, 2003:38).
- DDDs per inhabitant per year – this estimates the average number of days an inhabitant is treated per year (WHO, 2003:38).

Another unit of measurement used in drug utilisation studies is the prescribed daily dose (PDD). The WHO (2003:39) defines the PDD as "the average dose prescribed according to a representative sample of prescriptions." In short it can be defined as the average dose prescribed per day (Cosentino et al., 2000:513). This unit of measurement illustrates the average amount of drug that is prescribed per day.

The PDD is closely related to the diagnosis of the illness that is being treated. This is important because the same drug can be used for two different illnesses and two different doses. For example, for anti-infective drugs such as antibiotics, the dose may increase with an increase in the severity of the infection (WHO, 2003:39).

The PDD is used in this study as a unit of measurement as the exact doses that were prescribed can be extracted from the database and therefore an average PDD can be calculated. The following equation was used in the database to calculate the PDD:

$$Daily\ dosage = \frac{Strength \times Quantitiy}{Days\ supplied}$$

3.3.5.3 Significance ratings

According to Tatro (2011), not all drug interactions are equally severe. By adjusting the dose or altering the time of administration, the side-effects of these drug interactions can be avoided. Drug interactions are assigned a significance rating in an ascending order of 1 to 5 according to the severity and documentation level of the DDI (refer to section 2.3.3).

3.3.6 DATA ANALYSIS

The data analysis consisted of several parts. In the first part all the data are analysed in order to collect information on the overall prescribing patterns of medicine items. The next part of data analysis consists of the extraction of all warfarin products from the total study population and then analysis of these products in order to collect information on the different prescribing patterns of warfarin products. In the next part of the data analysis all the medicine items that were prescribed with warfarin products are extracted from the database in order to analyse their prescribing patterns in conjunction with that of warfarin. The data analysis is done by means of the statistical analysis system, SAS (SAS Institute Inc., 2006).

Figure 3.1 is a graphical representation of how the data were analysed in this study.
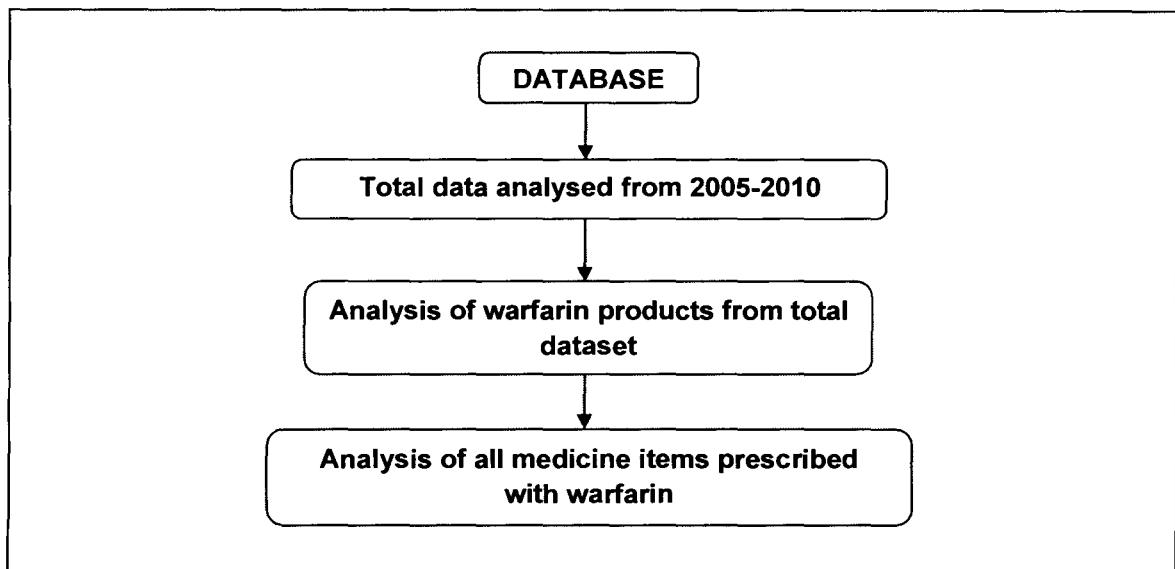
Figure 3.1: Graphical representation of the data analysis process.

3.3.7 STATISTICAL ANALYSIS

The following descriptive statistical analysis concepts were used to analyse the data according to the measuring instruments as discussed in section 3.3.3.

Before these descriptive statistical analysis concepts can be discussed, one point must first be made clear. All the equations discussed below can be calculated for either a sample or the whole population in a dataset. The researcher should take care to use the right symbols for the right equation. The symbols used for these calculations differ between sample and population measures. Table 3.1, as adapted from De Muth (1999:86), indicates the different symbols used for the sample and population measures.

Table 3.1: Symbols used in the calculations of samples and population measures

	SAMPLE	POPULATION
Mean	$\bar{X}$	μ
Standard error of the mean	$S_{\bar{x}}$	$\sigma_{\bar{x}}$
Standard deviation	S	σ
Number of observations	n	N

3.3.7.1 Average value (mean)

In order to determine the distribution of data in a particular data set, several descriptive measures are used to describe the distribution of data in a data set. The descriptive measure that is generally used is the mean, otherwise known as the average. The mean is calculated by adding up all the individual values occurring in the data set and dividing this value by the total number of the individual values present in the data set (De Muth, 1999:86; Waning & Montagne, 2001:83). To simplify, the mean is usually calculated as follows:

$$\mu = \frac{x_1 + x_2 + x_3 + x_n}{N}$$

μ = the mean or average

$x = x$ is the individual value in the data set and the subscript indicates the *i*th observation in the data set

N = number of observations in a population

In order to simplify this equation even more, the use of the summation sign (Σ) is incorporated to indicate the sum of all variables. The mean can also be referred to as the arithmetic mean (De Muth, 1999:87). The equation can thus be written as follows:

$$\mu = \frac{\Sigma x_i}{N}$$

μ = the mean or average

Σ = the sum of all variables

x_i = the individual value

N = number of observations in a population

3.3.7.2 Standard deviation

The standard deviation is known as a measure of dispersal. Measures of dispersion are used to define the variability in a given sample or population. The standard deviation defines the variability that occurs around the mean. The larger the value of the standard deviation the more variability occurs in the sample or population, and vice versa (Cohen & Lea, 2004:14; De Muth, 1999:91; Waning & Montagne, 2001:85). The equation for the standard deviation can be written as follows:

$$S = \sqrt{\frac{\Sigma(x_i - \bar{X})^2}{n - 1}}$$

S = the standard deviation

Σ = the sum of all variables

x_i= the individual value

$\bar{X}$ = the mean or average

n = number of observations in a sample

3.3.7.3 Confidence interval

Before the confidence interval can be discussed, attention must first be given to the standard error of the mean. The standard deviation depicts the variability that can occur within a sample around the mean value. The standard error of the mean depicts the possible variability that can occur within the mean itself. As with the standard deviation, the larger the value for the standard error of the mean, the more variability occurs within the mean (De Muth, 1999:128; Waning & Montagne, 2001:86). The standard error of the mean is calculated as follows:

$$S_{\bar{x}} = \frac{S}{\sqrt{n}}$$

$S_{\bar{x}}$ = the standard error of the mean

S = the standard deviation

n = number of observations in a sample

The confidence interval is a measure of the confidence in a certain estimate. The confidence interval depicts the range of values within an estimate, and allows for this depiction to be made with a certain degree of certainty. One can thus say with a certain degree of confidence that the estimated value lies in a certain range (Waning & Montagne, 2001:87, 88). The range used in confidence interval calculations can be obtained by means of the z-score. In any distribution of data the z-score depicts the number of standard deviations away from the mean. The value of the z-score illustrates the area of the curve that lies between the mean (z=0) and the distribution point above or below the mean. This means that if the z-score has a value of 1 and the mean is 0, and then +1 and -1 would be the points of standard deviation above and below the mean respectively (De Muth, 1999:119). The z-score can be calculated as follows:

$$z = \frac{X - \bar{x}}{S}$$

z = the z-score

X = the individual value

$\bar{x}$ = the sample mean

S = the standard deviation

Finally, the confidence interval can be calculated using the following equation:

$$CI_\mu = \bar{x} \pm z S_{\bar{x}}$$

CI_μ = confidence interval for a population mean

$\bar{x}$ = the sample mean

z = the z-score

$S_{\bar{x}}$ = the standard error of the mean

3.3.7.4 Cohens's *d*-value

In most cases when conducting a study, a null hypothesis must be generated. In the case when two groups must be compared, the null hypothesis usually states that there is no difference between the two groups. In order to reject the null hypothesis there must be a difference between the two groups (Steyn, 2012; Cohen, 1988:9).

A statistical instrument used to determine the statistical significance of the change that occurs between the two groups is Cohen's d-value. With Cohen's d-value, the difference between the two means of the groups being compared is divided by the largest standard deviation of the two groups (Steyn, 2012). This results in the following equation:

$$d = \frac{\overline{x}_a - \overline{x}_b}{S_{max}}$$

d = effect size (Cohen's d-value)

$\overline{x}_a$ = average value of group a

$\overline{x}_b$ = average value of group b

S_{max} = the maximum standard deviation between group a and b

The effect size of the d-value can be interpreted as follows (Steyn, 2012):

d-value = 0.1 small effect, only 1% of the variance of x can be explained by y, therefore there is no practical significance.

d-value = 0.3 medium effect, the relationship is observable by inspection, therefore it could be practically significant.

d-value = 0.5 large effect, there is a clear linear relationship that is practically significant.

3.4 RELIABILITY AND VALIDITY OF THE RESEARCH INSTRUMENTS

The data for analysis were directly obtained from the medicine claims database of the pharmaceutical benefit management company. Direct manipulation of the data by the researcher was therefore impossible. The research was done under the impression that all data obtained from the databases were precise and correct.

3.5 ETHICAL CONSIDERATIONS

A retrospective drug utilisation review was done on data provided by the database in South Africa with permission from the pharmaceutical benefit management company. The confidentiality of the patients was ensured by a numbering system. A unique number is assigned to every drug item. This number is used to identify the patient, pharmacy, medical practice or medical scheme through which the drug has gone. In this way confidentiality is sustained throughout the study.

Within the medicine claims databases the individual patients and prescribers could not be identified and the study was concluded anonymously. Ethical approval was obtained from the North-West University (NWU-0046-08S5).

3.6 DISCUSSION OF THE RESULTS OF THE EMPIRICAL INVESTIGATION

The discussion of the results of the empirical study will be included in Chapter 4.

3.7 CONCLUSION AND RECOMMENDATIONS

The conclusions and recommendations based on the results of the literature review and the empirical investigation will be discussed in Chapter 5.

3.8 CHAPTER SUMMARY

The purpose of this chapter was to explain the methodology of the empirical study. This chapter contained the general and specific research objectives of both the literature review and the empirical investigation. Other topics of discussion included the research design, the database, study population, classification systems, measurements and the statistical analysis. Chapter 4 will include the results of the empirical investigation.