STATISTICAL TECHNIQUES FOR IMPROVING THE REPEATABILITY OF AUTOMATED ECG INTERPRETATION.

Stephanie C. McLaughlin, B. Sc.

Thesis submitted for the degree of Ph.D.

To: The Faculty of Medicine
University of Glasgow

The Research described in this thesis was carried out in the University Department of Medical Cardiology, Royal Infirmary, Glasgow.

Stephanie C. McLaughlin

Signed:

Date:

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

TABLE OF CONTENTS

LIST OF TABLES

TABLE

TABLE

TABLE

TABLE

TABLE

LIST OF FIGURES

FIGURE

FIGURE

FIGURE

FIGURE

FIGURE

FIGURE

DECLARATION

I hereby certify that this thesis has been researched and written entirely by myself and that it has not been submitted previously for any degree.

Signed:

Date:

ACKNOWLEDGEMENT

I am deeply indebted to Professor P. W. Macfarlane for providing me with the opportunity of working in the field of electrocardiology and for all the genuine encouragement and assistance which he has enthusiastically given over the last few years.

Equally, I would like to express my thanks to Mr. T. C. Aitchison of the Dept. of Statistics for his unfailing support, guidance and friendship throughout the duration of this project.

This research would not have been possible without the help of many people. I would particularly like to thank Mr. B. Devine for all his patient programming assistance and Dr. F. Huwez for his helpful advice on various clinical aspects of the work. In addition, I am grateful to Mrs. C. Fullarton, Miss J. Kennedy, Mrs. K. McLaren and Mrs. M. Sneddon for their technical expertise in recording the numerous ECGs which were required.

I would also like to thank all my friends in the Departments of Medical Cardiology and Statistics and outwith the University for their helpful advice and assistance over the past few years.

Finally I wish to thank my family for the encouragement and support that they have given throughout my career.

SUMMARY

The electrocardiogram is a recording of the electrical activity of the heart. By convention, 12 separate leads are studied from which a diagnosis is made. This process of interpretation is a skill acquired by cardiologists over a period of years. During the past twenty-five years, computer assisted techniques have been developed to undertake such interpretations. It has been shown that if consecutive ECGs are recorded on a patient within several minutes, the computer diagnoses may differ because they are made independently of one another although all conditions remain unchanged. This discrepancy occurs when small changes in ECG measurements, from one recording to another, cause threshold values within the diagnostic program to be crossed, thereby producing conflicting diagnoses. The primary aim of the study described in this thesis was to develop techniques which would minimise such problems, thereby enhancing the repeatability of the ECG program developed in the Department of Medical Cardiology at the Royal Infirmary in Glasgow, whilst maintaining the heuristic framework of the diagnostic logic.

From a statistical perspective, the problem of lack of repeatability was tackled in the following ways. Firstly, a new approach to defining upper limits of normal ECG measurements was adopted. Conventionally, the Glasgow Royal Infirmary program categorises normal limits with respect to age and sex. These limits are discontinuous in nature and can contribute to a lack of repeatability in interpretation between consecutive recordings particularly when an individual's age-category has altered between visits.

These limits have been replaced with continuous equations which were calculated on the basis of a sample of 1338 'normals' using simple linear regression techniques. It was thought that the use of such equations, which change smoothly and continuously throughout the age range, would alleviate the problem of subtle differences in ECG readings from recording to recording producing inconsistent diagnoses.

The second stage of the problem was to consider how to deal with discrete thresholds between normal and abnormal, whether such boundaries were continuous or not. Many of the diagnostic decisions throughout the program have, until now, been determined by whether the observed value of a particular ECG measurement has attained a specified discrete threshold. These decisions can be regarded as score functions for ECG measurements which take the value K if the threshold is attained, and 0 otherwise. No account has been taken of the proximity between the measurement and the boundary value. This 'all or nothing' strategy has meant that an individual whose observed ECG measurement lies very close to, but below, a threshold value on one occasion and equally close to, but above, the same threshold value on a subsequent visit may receive two conflicting diagnoses.

A method was developed to replace these discrete thresholds with continuous functions, which take into account the natural day-to-day variation occurring in each ECG measurement, and to assign a new smoothed diagnostic score accordingly. Using the discrete score function as a basis, a smooth alternative was developed which increased gradually from 0 to a maximum of K. This smooth version of the scoring function was based on the family of cumulative distribution functions of the logistic distribution, suitably scaled to give

the desired maximum score. In addition, the steepness of the new smoothed step was dictated by the amount of day-to-day variation in the particular ECG measurement under consideration.

Often, more than one criterion needed to be satisfied before appropriate action could be taken, so an algebra which would take account of combinations of criteria was required. Some basic mathematical rules such as the intersection rule and the union rule were used as building blocks upon which to construct such a system.

A database of 590 ECG recordings obtained on consecutive days from 295 patients (using no standardising procedures) was collected in order to estimate the amount of 'normal' day-to-day variation in the relevant ECG variables.

Preliminary investigations were undertaken with the aid of some simple exploratory plots of the data. These plots indicated that two models for day-to-day variation were required, each accounting for the fact that the amount of variation may or may not be dependent on the magnitude of the ECG measurement of interest. In most cases, the simpler of the two models was adequate, although in certain situations a slightly more complex model was required. Estimates of day-to-day variation were obtained for all the relevant ECG variables and subsequently used in the calculation of the new smoothed scores.

The methods were then applied to examples from three major diagnostic sections of the program. The diagnosis of left ventricular hypertrophy (LVH) was used to represent the section of the program dealing with Hypertrophy, while the repeatability of the diagnosis of Myocardial Infarction (MI) was assessed on the basis of the agreement in the identification of inferior MI (IMI). Similarly, the diagnosis of

inferior ST depression was used as an indicator of the repeatability of the section of the program dealing with ST changes.

Repeatability was assessed in three ways:

1) on the basis of 660 ECGs which were recorded on two consecutive days from 330 patients;
2) using ECGs which were recorded within the space of a few minutes (without removing and subsequently replacing the electrodes) from 249 patients; and
3) using an artificial method of splitting the 330 day 1 ECG tracings into 2 digital representations.

In terms of overall repeatability, it was discovered that of the 330 pairs of repeat ECGs, 266 (81%) interpretations were completely in agreement with respect to the three previously mentioned diagnoses when the conventional version of the program was used. The corresponding numbers for the 249 pairs of minute-to-minute ECGs and 330 split ECGs were 222 (89%) and 304 (92%) respectively. Adopting the smoothing techniques in the conventional program increased the number of pairs of ECGs which were entirely in agreement in all three cases. Of the 330 pairs of day-to-day ECGs, 291 (88%) were in agreement, as were 236 (95%) and 315 (95%) of the minute-to-minute and split ECGs. There have therefore been substantial reductions in the percentages of pairs of ECGs exhibiting discrepancies. For example, the percentage of minute-to-minute ECG recordings producing inconsistent diagnoses has been reduced from 11% to just 5% which represents a 55% overall reduction in error. There were similar improvements for both day-to-day and split ECG recordings (37% and 38% overall reductions in error respectively).

When considering Type A statements relating to LVH and IMI only (i.e. statements referring to clinical conditions which can be

verified via non-electrocardiographic means such as echocardiography, serum enzyme levels etc.), it was demonstrated that the conventional Glasgow program was perfectly repeatable in 295 of the 330 pairs of the day-to-day ECG recordings (89%), in 232 of the 249 pairs of ECGs which were recorded within the space of a few minutes (93%) and in 96% of the artificially split ECGs (i.e. 317 out of 330 tracings). Implementation of the smoothing techniques resulted in improved repeatability in all three situations. For the day-to-day ECGs, the repeatability rose to 95% and the level of agreement from recording to recording was 98% for both the minute-to-minute ECGs and the artificially split ECGs, proving that the methods developed in this thesis are of benefit, even in situations when the level of agreement from recording to recording is reasonably acceptable.

The Glasgow program is long-established and has been associated with an acceptable diagnostic accuracy. In this respect, a study of the sensitivity and specificity of the new techniques compared to the old showed no significant difference in a group of 84 patients with LVH diagnosed by M-mode echocardiography.

There has clearly been a significant and important increase in the reliability of the ECG program as a result of implementing the methods which have been developed. Therefore this thesis has demonstrated that the use of statistically based techniques for discriminating between normal and abnormal has enhanced the repeatability of computer assisted reporting of electrocardiograms.

CHAPTER ONE: ELECTROCARDIOGRAPHY.

1.1 INTRODUCTION.

Electrocardiography is the term used to describe the process which allows the electrical activity of the heart to be displayed in graphical form. The first human electrocardiogram (ECG) was recorded approximately one hundred years ago by Waller (1887) and since then, advances in modern technology have enhanced the field of electrocardiography and promoted its uses world-wide. Increasing popularity and demand for recordings of electrocardiograms led to research into the possible role of the digital computer as an analytic tool and the early 1960s saw initial automated analyses for both the 3-orthogonal lead and the 12 lead ECG (Pipberger et. al. 1960, Caceres et. al., 1962; Caceres, 1963; Stallmann et. al, 1961; Klingeman and Pipberger, 1967). Many benefits were anticipated, notably:

(i) the ability of computers to deal with the rapidly expanding demand for ECG recordings,

(ii) increased accuracy in the measurement of wave magnitudes and durations, and

(iii) the probable reduction of intra-observer and inter-observer variation.

1.2 ANATOMY

The heart is a muscular cone-shaped organ and is composed almost entirely of cardiac muscle (myocardium). It is divided into four chambers - the left atrium, the right atrium and the left and right ventricles. The upper chambers - the atria - receive blood from the veins, while the lower chambers - the ventricles - propel blood into the arteries (see Figure 1.1).

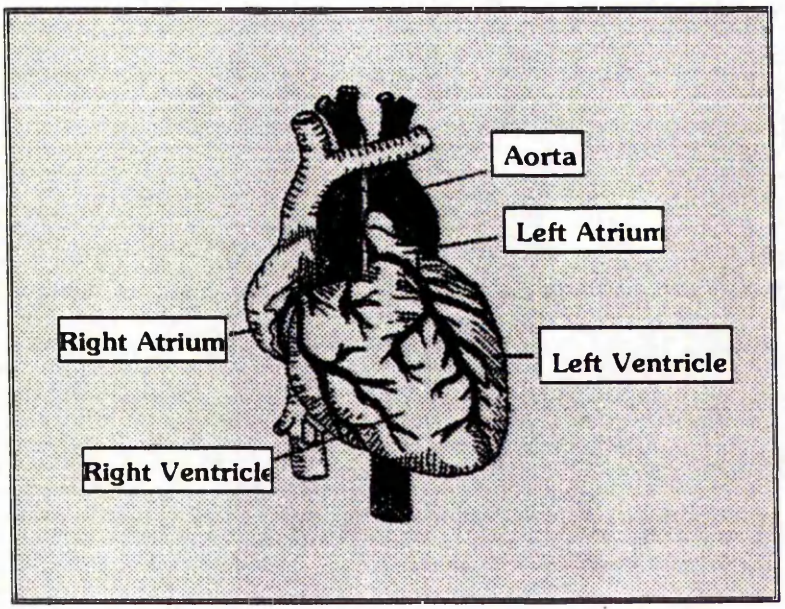


Figure 1.1 **The heart**

There are three types of cell present which work together to create and conduct the impulses which result in the contraction of the heart. Electrical impulses originate from the pacemaker cells which are located in the sino-atrial node. Conducting cells then transmit these impulses to the atrio-ventricular node via atrial internodal pathways and these impulses then travel to the ventricles via the

bundle of His and the left and right bundle branches. Conduction then spreads through the specialised tissue in the ventricles known as 'Purkinje fibres' and into cardiac muscle itself. Stimulation of the muscle cells causes the mechanical contraction which is associated with each heart beat.

1.3 THE FIRST ECG RECORDING

Contraction of the heart is associated with changes in polarity of the electrical charges on the surface of the myocardial cells. In the resting state, the cells are positively charged and when the cells are excited they enter a state of physical activity. This state comprises the electrical process known as depolarisation which changes the positive charge to a negative charge. Subsequent repolarisation then returns the cells to their resting state.

The electrical forces produced during the processes of depolarisation and repolarisation initiate and maintain the beating of the heart and the resulting electrical signals are transferred to the skin through electrically conductive tissue. Thus, potential differences can be recorded in the form of an electrocardiogram by attaching electrodes to the body's surface. Much research which took place in the mid-nineteenth century helped pave the way for Waller (1887) who recorded the first known human ECG in 1887. Galvani probably set the scene by exploring current flow in frogs with the help of his galvanometer which was able to detect small electrical signals. This led the way for other researchers such as Marey (1876) who photographically recorded the electrical activity of a frog's heart using

Lippmann's electrometer and Burdon-Sanderson and Page (1878) who investigated the electrical activity of the tortoise heart.

Willem Einthoven developed a much more sensitive string galvanometer (Einthoven, 1901) which was designed with the purpose of recording the electrical activity of the heart. The electrical activity was detected by the quartz string in the galvanometer and the string's movement could be recorded photographically to make a tracing.

The ECG takes the shape of a waveform displaying characteristic peaks and troughs. Einthoven introduced the PQRST terminology to describe the deflections of the ECG waveform as illustrated below.

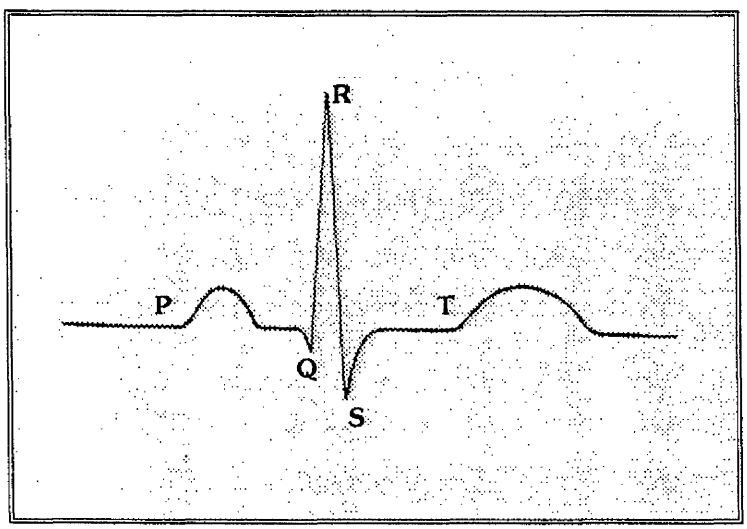


Figure 1.2 **The ECG waveform**

Each section of the waveform is produced by a different aspect of the electrical activity of the heart. The components may be summarised briefly as follows:

P wave	-	atrial depolarisation
QRS complex	-	ventricular depolarisation
ST segment	-	ventricular repolarisation (phase 2)
T wave	-	ventricular repolarisation (phase 3)

The P wave is of a smaller magnitude than the QRS complex since the atrial mass is less than the ventricular mass and hence repolarisation of the atria produces a relatively small deflection which by virtue of its timing is normally overwhelmed by the larger QRS complex.

Several factors determine the magnitude and the direction of the deflections in the ECG, the most influential being the location of the electrodes on the body surface and the direction of the cardiac impulses in relation to the measuring system. By convention, if the depolarisation impulse travels towards the positive electrode an upward deflection will be recorded and vice versa.

1.4 LEAD SYSTEMS

There are two commonly used methods of displaying the electrical forces in the heart. Waller had first introduced a single dipole (1889) to represent the cardiac electrical activity and Einthoven and colleagues postulated that the mean electrical axis of the heart could be represented by a vector (Einthoven, Fahr, de Waart, 1913). The first attempt to form a vectorcardiogram (VCG), i.e. a loop that traced the direction of the resultant vector throughout the cardiac cycle, was published in an early paper by Williams (1914) who derived the cardiac vector from Einthoven's standard bipolar limb

leads. Through the following years, vectorcardiography evolved and Frank (1956) developed a 'corrected orthogonal-lead system' which essentially recorded three leads X, Y and Z measuring the component of the resultant cardiac electrical force in three mutually perpendicular directions. This is the most popular lead system for vectorcardiography and is still in use in various centres today.

Bipolar leads were used by Einthoven et. al. (1913) to measure the potential difference between two limbs at points remote from the heart, hence the term bipolar limb lead. Electrodes were attached to the left and right wrists and to the left ankle to make the following connections :

Lead I	Left Arm	→	Right Arm
Lead II	Left Leg	→	Right Arm
Lead III	Left Leg	→	Left Arm

This may be represented mathematically as

$$I = E_L - E_R$$
$$II = E_F - E_R$$
$$III = E_F - E_L$$

where E_L, E_R and E_F signify the potential at the left arm, right arm and left leg, and I is the potential in lead I etc.

It follows that, at any instant in the cardiac cycle,

$$I + III = II$$

and this is known as Einthoven's Law (Einthoven et. al, 1913). In a practical sense this represents an important observation since it

effectively renders one of the leads redundant, allowing it to be calculated from knowledge of the other two.

Assuming that the right arm, left arm and left leg are arranged symmetrically around the heart and that the thorax represents a homogeneous conductor, the limb leads denoted I, II and III form an equilateral triangle with the heart lying approximately at the centre. This is known as Einthoven's triangle (Einthoven et. al., 1913).

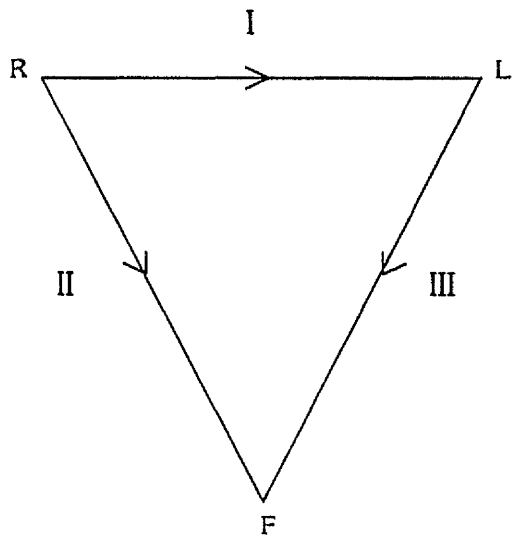


Figure 1.3 **Einthoven's triangle**

Investigation of the inhomogeneity of the thorax led Burger and an associate to redefine this triangle and to construct a scalene triangle - Burger's triangle - of which Einthoven's triangle is a special case (Burger and van Milaan, 1946).

1908 saw the commercialisation of Einthoven's electrocardiograph by the Cambridge Instrument Company (Macfarlane, 1989a) and its potential value was rapidly recognised by the rest of Europe and the United States.

Continuing research and collaboration resulted in Wilson's central terminal (Wilson et al, 1934). Electrodes were attached to the right arm, left arm and left leg as before and these were then averaged to form the reference or central terminal. This allowed an 'exploring electrode' to be used to record potential variation at each of the limbs as follows:

$$VR = E_R - E_{WCT}$$

$$VL = E_L - E_{WCT}$$

$$VF = E_F - E_{WCT}$$

$$E_{WCT} = \frac{1}{3}(E_R + E_L + E_F)$$

where E_R, E_L and E_F are as before, E_{WCT} denotes the potential at the central terminal which is relatively constant and VR, VL and VF denote the potentials measured at each of the limbs (Macfarlane, 1989b). Since the recorded potential effectively reflected potential variation at a single point because E_{WCT} is essentially constant, these leads were termed unipolar limb leads. When recorded in this way, the deflections were inconveniently small. However, slight modifications made by Goldberger (1942) to Wilson's central terminal produced augmented unipolar limb leads - denoted aVR, aVL and aVF - which essentially increase the voltages recorded by the unipolar limb leads by 50%, i.e.

$$aVR = \frac{3}{2}VR$$

$$aVL = \frac{3}{2}VL$$

$$aVF = \frac{3}{2}VF$$

(It can be shown that, at any instant, aVR + aVL+ aVF = 0).

Wilson's central terminal allowed a further six leads which recorded potential variation at single points on the chest to be introduced. These leads were called unipolar chest leads for obvious reasons and were initially denoted V1, V2, V3, V4, V5 and VE. Leads V1 to V5 were derived from electrodes placed at designated areas on the chest and lead VE was positioned at the tip of the ensiform process (Kossmann and Johnston, 1935). Later, a committee of the American Heart Association published recommendations which were agreed upon by the Cardiac Society of Great Britain and Ireland on the positioning of six precordial leads V1, V2, V3, V4, V5 and V6 (Committee of the American Heart Association, 1938).

The combination of I, II, III, aVR, aVL, aVF, V1, V2, V3, V4, V5 and V6 forms the conventional 12-lead ECG which is the standard method for recording electrocardiograms virtually world-wide.

A version of the waveform described previously is found in each of the 12 leads and the resulting ECG may therefore contain a vast amount of information. The ECG is usually interpreted in conjunction with any historical and clinical data which may be available.

1.5 DIGITAL COMPUTERS

Electrocardiography has become one of the most popular clinical tests in hospital practice and the subsequent demand has resulted in research into the possible role of the digital computer as an analytic tool not only for obtaining and manipulating the ECG, but also for interpretative purposes (Macfarlane, 1974). Research into automated

analysis of both the 12 lead and the 3-orthogonal lead ECG began in the late 1950s and rapid progress was being made by the early 1960s (Pipberger et. al, 1960, Caceres et. al., 1962, Caceres, 1963; Stallman et. al. 1961; Klingeman and Pipberger, 1967). These advances provided a stimulus for several other research groups (Bonner et. al., 1972; Macfarlane, Lorimer and Lawrie, 1971) and by the early 1970s automated methods being used by several teams yielded results which were in reasonable agreement with interpretations made by physicians. Such methods were implemented with the proviso that each automated ECG report be checked by medical staff before distribution. However, concern as to the repeatability of automated methods meant that further research was required.

1.6 DIAGNOSTIC PROGRAMS

Since the inception of automated electrocardiography two distinct methods for interpretation have been followed, each appealing to separate groups of researchers and benefiting from further developments. These methods are the statistical and the deterministic approaches to ECG classification, each having its own advantages and disadvantages.

Statistical classification methods are largely based on multivariate analysis and the early work in this particular area was done by Cady, Kimura and others (Cady et. al., 1961; Kimura, Mibukura and Miura, 1963). Pipberger has for long been a strong advocate of such techniques on the basis that they are more stable

and more accurate than their deterministic counterparts (e.g. Klingeman and Pipberger, 1967). Initial work was done to distinguish between two populations (Eddleman and Pipberger, 1971; Goldman and Pipberger, 1969). Eddlemann used linear discriminant-function analysis with 15 ECG variables to diagnose correctly 88% of an independent test set of autopsy cases of myocardial infarcts. Goldman and Pipberger (1969) applied statistical techniques to a group of patients with conduction defects.

Cornfield et. al. (1973) used the established methodology in the multi-group situation which considered seven possible diagnostic categories, namely normal, anterior myocardial infarction, posterior myocardial infarction, lateral myocardial infarction, left ventricular hypertrophy, right ventricular hypertrophy and pulmonary emphysema.

A Bayesian approach was adopted whereby the posterior probability $P(i|\underline{x})$ of an individual with ECG vector $\underline{x}$ belonging to diagnostic class i is calculated on the basis of the prior probability g_i of belonging to that class, i.e.

$$P(i|\underline{x}) = \frac{f(\underline{x}|i)g_i}{\sum_{j=1}^{m} f(\underline{x}|j)g_j}$$

where

m = the number of possible diagnostic categories

$f(\underline{x}|i)$ = the conditional probability of $\underline{x}$, given patient belongs to i.

More recently, the use of logistic classification models has also been assessed (Willems et. al., 1986).

Deterministic methods were initially established by Caceres (Caceres, 1963). Such methods are built largely on experience and expertise which has been accumulating over many years, resulting in a logical path of decision rules which imitate the role of the cardiologist.

The development of both deterministic and statistical methods requires a substantial amount of data from different populations since it is known that ECG parameters vary according to age, sex and race. For example, Pipberger's AVA program which has been based on a male and war-veteran population is unlikely to perform in the same way on a group of young women.

Small changes in measurements may alter diagnostic statements in deterministic programs since these techniques use measurements sequentially. Thus statistical techniques have been claimed by some to be more robust when compared with their deterministic counterparts. For example, Walston, Harley and Pipberger (1974) demonstrated that multivariate statistical techniques correctly identified mitral stenosis in 74% of the patients whom they studied, compared to 44% when a deterministic approach was used. Similar, though not so striking, results were provided by Brohet et. al. (1984) who obtained a correct classification rate of 85% for interpreting paediatric ECGs using a statistical program compared to 79% when a deterministic approach was used. In the late nineteen seventies, Willems (1977) demonstrated the higher accuracy of statistical programs by presenting results showing that the overall performance of the multivariate AVA program was 77.6% compared to 68.9% for version 7603 of the decision-logic based TNO system when attempting to diagnose myocardial infarction. Later, a program based on logistic classification models was shown to be more accurate than

the AVA program and the HP78 program, with overall performances being 80.3%, 72.8% and 63.3% respectively when diagnosing a variety of abnormalities (Willems et. al, 1986).

However, since multivariate techniques are typically based on the assumption that the primary diagnostic categories are mutually exclusive, the training set required to develop a statistical program with the ability to distinguish combinations of disease is likely to exceed the number of well-documented cases since each combination must be considered as a separate category (van Bemmel et. al., 1971). It is also apparent that the use of prior probabilities in statistical programs can influence the results. Indeed, the overall accuracy of the AVA program was decreased by 20% when prior probabilities were set equal for all possible categories (Pipberger et. al. 1975).

Deterministic methods appeal more to the end-user (i.e. the cardiologist) than statistical methods because they are more readily understood. Diagnostic criteria may also be selected on the basis of a knowledge of well-established electrophysiological processes. Deterministic methods are also reasonably flexible so that alterations to existing criteria can be implemented, and new diagnostic categories are easily added (Bailey and Horton, 1977; Kors and van Bemmel, 1990). Recently it has been demonstrated that, when cardiologist opinion is the gold standard deterministic methods have a higher accuracy than statistical techniques. However, when compared with the 'truth' (based on independent clinical evidence), statistical programs are more accurate (Willems et. al., 1991).

Fuzzy set theory has also been used in an attempt to interpret the ECG (Smets et. al, 1977, Degani and Bortolan, 1986). However,

this approach has not been widely used by clinicians and is therefore difficult to assess.

1.7 THE CSE DIAGNOSTIC STUDY

As the number of commercially available interpretative ECG systems increased, so did the need for the development of evaluation methods which would serve to assess the diagnostic performances of the various ECG programs. Thus an international project was founded - the CSE Diagnostic Study - which aimed to establish common standards in quantitative electrocardiography. It was anticipated that, along with the assessment of diagnostic performance, exchange of information, improved co-operation between investigators, development of quantitative test procedures and improvements in measurement precision would also be achieved (van Bemmel, 1986).

The first stage of the CSE project was to develop standards for ECG measurements by evaluating measurement variability with respect to a reference based on results provided by referees. The second stage, the diagnostic study, aimed to assess the diagnostic performance of commonly used ECG interpretative programs by comparing their interpretations with those provided by cardiologists and with the 'truth' which was established on the basis of ECG-independent evidence such as catheterization, echocardiography and physical examinations (Willems et. al, 1991).

A pilot study was agreed upon comprising 250 well-documented ECG recordings (obtained in Glasgow and Dublin) which belonged to one of seven groups - normal, left ventricular hypertrophy, right

ventricular hypertrophy, biventricular hypertrophy, anterior myocardial infarction, inferior myocardial infarction and combined infarction. Twelve different computer programs analysed the recordings - nine using the standard 12-lead ECG and three using the VCG. The recordings were also analysed independently by a panel of six cardiologists. The pilot study demonstrated several problems due to the different approaches used by the various programs (e.g. deterministic programs, statistical programs and programs based on fuzzy logic) and to the different terminology used by the centres. As a result, common CSE codes were introduced and each centre was required to apply a mapping scheme to produce the relevant code from the diagnostic statement. The performances of the diagnostic algorithms were compared using a statistical method outlined by Bailey et. al. (1988). Numerous preliminary results have been published (Willems et. al., 1987) but should be interpreted with caution given that they are based on a limited amount of data (Willems, 1988).

The final database consisted of 1220 well-documented and clinically validated ECGs representing the seven diagnostic categories previously mentioned. Individual program and cardiologist results were compared with the 'truth' and each program was also compared with the combined interpretations of eight cardiologists. Results have been published (Willems et. al., 1991) in conjunction with other relevant information which concludes that the CSE Diagnostic Study has been a valuable exercise and that improvements can still be made to further the field of computerised electrocardiography.

1.8 THE GLASGOW PROGRAM

Since developments made throughout this thesis will be based around the structure of the Glasgow Program it is of relevance to describe the Glasgow approach briefly. Full details have been published elsewhere (Macfarlane et. al., 1990), while there will be an expanded discussion on relevant diagnostic criteria in subsequent chapters of this thesis.

The current 12-lead ECG program originated in 1977 after several years of developing and applying techniques for analysis of the 3-orthogonal lead ECG. The reason for this change was that mini computers became faster in the late nineteen seventies and the availability of microprocessors meant that multiple leads could be recorded simultaneously. Macfarlane et. al. (1980) developed a hybrid system around that time which allowed information from the 3-lead and the 12-lead ECGs to be combined.

ECGs are recorded either by a locally designed and built electrocardiograph (Watts and Shoat, 1987) or a MINGOREC 4 (Siemens-Elema), both acquiring ECG leads simultaneously and digitising them at 500 samples per second for input to a central computer. The Glasgow program is now implemented on a MICROVAX computer which uses the operating system ULTRIX-32 and broadband links allow the direct transfer of the ECGs from the locally designed recording carts to the computer. The advent of microprocessor technology in the 1970s means that this central ECG system is now more accessible to clinics, health centres and smaller hospitals since the ECG data can also be transferred from newer electrocardiographs via modems and telephone lines.

The interpretative section of the program is deterministic in nature and considers all usual cardiac abnormalities. There are, for example, sections dealing with Conduction Defects, Ventricular Hypertrophy, Myocardial Infarction and ST-T abnormalities. The deterministic approach has been adopted largely on the basis of the disadvantages of the statistical approach previously described. Statistical programs are typically based on the assumption that the primary diagnostic categories are mutually exclusive and often this is not the case. Indeed, from over 30,000 ECGs which are recorded in Glasgow Royal Infirmary each year, there may be innumerable combinations of diagnostic statements. The training set for a statistical program which will allow identification of many combinations of cardiac abnormality is impractical since vast amounts of data representing each category will be required in addition to data representing each subdivision of the population (sex, age, race). Furthermore, manipulation of prior probabilities can often lead to misleading diagnostic results. Finally, clinical information may be used more easily in deterministic programs and the output from such programs is more acceptable to clinicians.

1.9 SERIAL CHANGES

It is important to be able to examine the serial changes in an ECG over time since this will allow monitoring of the progression of a particular disease.

Methods for detecting serial changes in the 3-orthogonal lead ECG have been investigated and several different methods of comparing ECGs assessed (Macfarlane, Cawood and Lawrie, 1975).

Selected measurements pertaining to the detection of sequential changes in myocardial injury were stored and criteria produced. The analysis of day-to-day and beat-to-beat variation in a group of normal serial ECGs (Cawood et. al., 1974) provided knowledge of the amount of variability which could be construed as 'normal' and the criteria for the detection of sequential changes were thus established.

Currently, in the Glasgow program, three ECGs are compared for an individual in any given year (Macfarlane, 1989c). The primary record is the first recording and the other two are those most recent prior to the current ECG. Where ECGs are compared with earlier recordings there will be some indication of whether there have been significant changes or not.

1.10 DAY-TO-DAY VARIATION

Day-to-day variation of the ECG has been investigated by several groups. Cawood et. al. (1974) reported differences in QRS complex and ST-T segment data from day to day and from beat to beat which were then used as a guide for the detection of abnormal changes from one recording to the next using the Glasgow program. Willems and colleagues (Willems, Poblete and Pipberger, 1972) demonstrated significant repeat variation in many measurements which was subsequently reduced, but not eliminated, when electrode positions were marked with a coloured skin dye. The reproducibility of the IBM ECG program was examined (Tuinstra, 1986) and significant variation in diagnostic statements discovered. Machado et. al. (1991) investigated the repeatability of a commercial version of the Glasgow program by comparing consecutive ECGs which were obtained

without replacing the electrodes. Of the 405 pairs of ECGs which were considered, type A statements (i.e. statements referring to conditions which can be verified via non-electrocardiographic sources such as echocardiography and physical examination) were reproduced in 93% of cases. This reproducibility is superior to that obtained by human readers.

In 1974, Bailey and his associates suggested a method of testing the reproducibility in ECG program performance which was independent of the clinical accuracy of the program (Bailey et. al., 1974). Instead of using two consecutive recordings from each patient, two digital representations from the same tracing were obtained. Initially the analogue data sets were collected at 1000 samples per second from which two data sets were extracted representing the same analogue data only this time digitised at 500 samples per second and separated in time by one msec. The performance of four programs was assessed and results reported. The IBM (1971) program was superior both to version D of the PHS program and to the Mayo Clinic program of 1986 exhibiting identical diagnostic statements in 76% of tracings compared to 43.3% and 60% for the PHS(D) and Mayo Clinic programs which were available at that time. Analogue filtering had the effect of increasing the percentages of identical statements to 79.7% and 49.8% for the IBM and PHS(D) programs respectively. The fourth program which was examined in this way was version 3.4 of Pipberger's automatic vectorcardiographic analysis (AVA) program (Bailey, Horton and Iscoitz, 1976). Identical readings were observed in 82.4% of 217 filtered ECGs.

Lack of repeatability remains a cause for concern and there may be many contributory sources of such variation. Crucial to the

reproducibility of any diagnostic program is the reproducibility of the ECG (Tuinstra, 1986) and noise and measurement errors may mean that deterministic programs may be more susceptible to repeat variation than their statistical counterparts (Willems, 1977). Bailey and his associates stressed the importance of reproducibility testing of ECG interpretative systems (Bailey et. al, 1976) and this remains a fundamental issue when assessing and evaluating diagnostic programs.

In the following chapter the possible sources of day-to-day variation will be considered whilst the subsequent two chapters will attempt to estimate the normal amount of day-to-day variability associated with ECG measurements. It is anticipated that once accurate and reliable estimates of normal variation have been provided for relevant ECG parameters, the extent to which a particular criterion allows for this variability may be assessed. To improve on this, smoothing techniques will be described in this thesis and applied to the deterministic program which is currently being used in Glasgow with the primary aim of improving repeatability.

CHAPTER TWO:

DIFFERENCES IN ECG WAVEFORMS.

2.1 INTRODUCTION.

Despite the many advantages gained through automated interpretation of the electrocardiogram, there remains cause for concern about the lack of repeatability of such techniques. By adopting a computerised approach to the analysis of ECGs, certain sources of errors have been minimised or indeed eliminated. For example, human error is removed and accuracy in wave measurements increased (Macfarlane and Lawrie, 1974). However, it can happen that on two consecutive clinic visits two conflicting ECG diagnoses may occur when there has been no clinically significant change in the pattern of the ECG waveform. The reasons for this are plentiful.

Before considering possible causes of repeat variation it is perhaps of value to mention several factors which can influence the appearance of the ECG from one individual to another.

2.2 DIFFERENCES IN ECG APPEARANCES.

Many ECG measurements have been found to vary with sex, race and age. Several other variables have also been found to correlate with ECG parameters but in the main, once these three important

factors have been taken into consideration, measurements such as height and weight provide little additional information on variability.

2.2.1 Differences due to sex.

Young adulthood appears to be the period in which sex differences in the ECG become obvious, such differences perhaps being due to the higher fat content of females and the presence of breast tissue, resulting in lower voltage amplitudes (Macfarlane and Lawrie, 1989d). Differences in the SV2 amplitude between male and female Caucasians are illustrated in Figure 2.1.

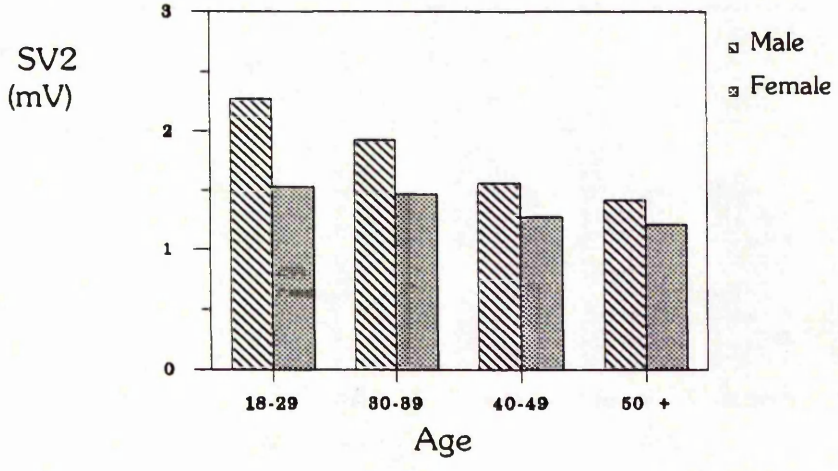


Figure 2.1 Distribution of the mean SV2 amplitude vs. Age (based on 719 males and 584 females)

2.2.2 Differences due to race.

Differences in the magnitude of ECG variables can also be observed when comparing races. The following histogram (Figure 2.2) illustrates how the amplitude of the S wave in lead V2 varies between Caucasians and Chinese (Chen, Chiang and Macfarlane, 1989).

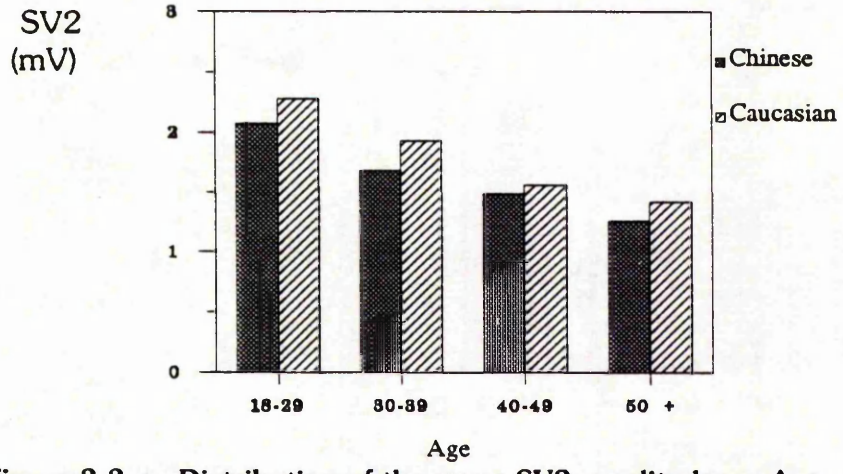


Figure 2.2 **Distribution of the mean SV2 amplitude vs. Age (Based on 719 male Caucasians and 205 male Chinese)**

2.2.3 Differences due to age.

Age is certainly the most influential factor to consider when interpreting ECGs. While sex perhaps does not influence the ECG substantially in infancy and old age, the importance of age can be demonstrated from birth until death and it is for this reason that age is used in many areas of diagnostic programs. Figure 2.3 demonstrates how the mean SV2 amplitude varies from young adulthood onwards for both males and females.

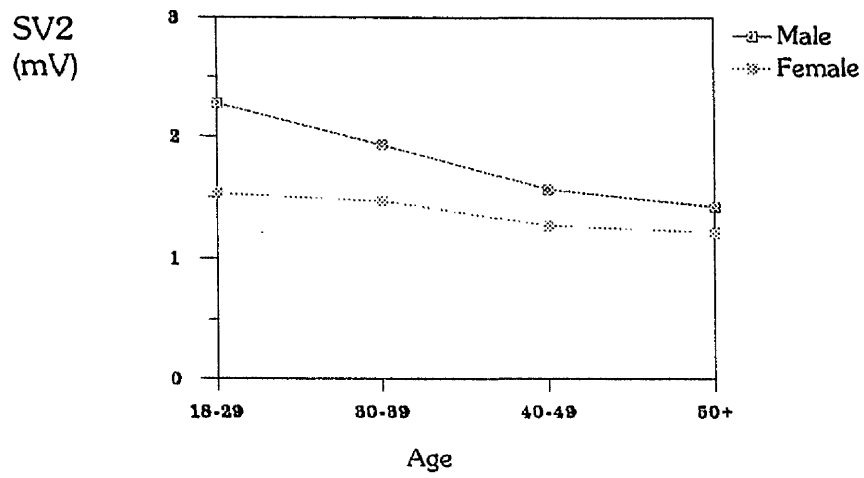


Figure 2.3 Mean SV2 amplitude vs. Age (based on 719 males and 584 females)

Sex and race are used as categorical variables in deterministic programs. This means that they are separated into categories for diagnostic decision making, normally consisting of a small number of classes (i.e. two for sex and two for race when considering Caucasian and Chinese only). Age, however, is a continuous variable which until now has been categorised into several mutually exclusive groups by all deterministic programs, e.g.

18-29, 30-39, 40-49, 50+

Upper limits of normality for ECG variables are usually constructed for each age group, stratified by sex and race, resulting in several distinct categories into one of which a particular individual will belong.

The above categorisation of age may cause problems if, between consecutive recordings, an age threshold has been crossed. Different upper limits of normality apply to the magnitude of the S wave in lead V2 for adult males under 30 years than for those aged between 30 and 40 so that if a particular individual's SV2 measurement has not changed from one clinic visit to the next but his 30th birthday

has occurred in the interim, it is possible that his ECG may be considered as normal on one occasion and abnormal on the next. This is one contributory factor to the problem of repeat variation and arises as a result of the discontinuous nature of many boundaries.

It is clearly unlikely that an individual's ECG waveform will be completely identical from one recording to the next. There will typically be small, clinically insignificant changes in wave amplitudes and durations. Regardless of how minute such changes may be, conflicting computer diagnoses can occur if discrete boundaries are crossed and this phenomenon, combined with the age-related discontinuities, forms the crux of the problem of repeat variation.

2.3 DAY-TO-DAY VARIATION.

Variation in the PQRST amplitudes and durations can potentially lead to a lack of repeatability in computer diagnoses from one ECG recording to another and this naturally gives cause for concern. Knowledge of the causes of day-to-day variation allows an experienced cardiologist to report no significant change between consecutive recordings although a method by which such variation may be quantified would prove appealing. On the other hand, cardiologists are not infallible. The CSE diagnostic study found that the median repeatability of cardiologists when given the same ECG on two separate occasions was 81.8% (Willems et. al., 1991).

Lewis recognised that many different ECG patterns could be observed in a group of healthy individuals although he largely assumed that the ECG for any given individual remained constant (Lewis and Gilder, 1912).

Simonson first addressed the question of day-to-day variation and emphasised the need for knowledge of such variation in order that serial electrocardiography be carried out (Simonson, Brozek and Keys, 1949). He suggested that eating a meal between recordings may alter the ECG significantly and provided measures of the amount of variation exhibited between eleven consecutive recordings taken over a period of two months on twelve normal young men.

Graybiel and colleagues demonstrated the effects of inhalation of tobacco smoke on the ECG (Graybiel, Starr and White, 1938). Drinking iced water between recordings has also been shown to alter ECG appearances (Sears and Manning, 1958).

Electrode placement is another cause of repeat variation since varying distances between the electrode and the heart will result in different ECG potentials and durations being recorded. It is therefore important that due care and attention is given to the placing of the electrodes when ECGs are being recorded.

Day-to-day variability can be reduced by the order of 25% when electrode positions are marked between recordings. Willems et. al. (1972) demonstrated such a reduction when comparing 3-orthogonal lead electrocardiograms which were recorded on individuals whose electrode positions were marked with a coloured skin marker and on those whose electrode positions were unmarked. Although repeat variation was reduced to a certain extent with such an approach, it was not eliminated.

Cawood et. al. (1974) provided estimates of day-to-day and beat-to-beat variation in normal 3-orthogonal lead ECGs and, given that the electrode positions were not marked, the variation proved quite substantial. Although such estimates may be reduced when electrode positions are marked, this is generally impractical in a busy hospital

and there still remains sufficient concern as to the other factors contributing to repeat variation.

2.4 SUMMARY.

The sources of variation in ECG appearances which have been described can be divided into two categories -

a) 'between-individual' variation and

b) 'within-individual' variation.

'Between-individual' variation will arise due to the fact that many ECG variables have been observed to vary with sex, race and age. Many ECG-analysis programs can deal with the former type of variation by stratifying by age, sex and race although using age as a discrete variable can give rise to a lack of repeatability.

'Within-individual' variation is more difficult to quantify since there are many different sources. Chapter 4 will attempt to provide, for each ECG variable of interest, an estimate of the day-to-day variation which will contain, as far as possible, all contributory factors. These estimates will then, in turn, be used in conjunction with a modification to the deterministic process currently in use in Glasgow (which will be outlined in a subsequent chapter) with a view to improving the repeatability of diagnoses.

CHAPTER THREE:

SOURCES OF REPEAT VARIABILITY IN THE ECG:

Age-categorised 'normal ranges'.

3.1 INTRODUCTION.

In the previous chapter attention was given to the fact that upper limits of normal for selected ECG variables are stratified by age, sex and race. The reason for this is that many ECG measurements differ across sub-groups of the population and it would be inadvisable to make use of the same criteria for young male Caucasians as are used for older female Chinese.

Having been stratified by race and sex, many limits are usually categorised by age and the problems arising from such a procedure have already been discussed. While these limits will remain stratified by race and sex, it is appealing to make use of the continuous nature of age in providing smoothed age-related upper limits of normality where appropriate.

3.2 BACKGROUND TO SAMPLE SELECTION.

The establishment of normal limits is made on the basis of a sample of assumed healthy individuals. Initially the target population (i.e. the population of interest) must be identified and from this, the study sample chosen. It is important that the study sample is representative of the target population since any inferences based on

the smaller group (which will be more accessible than the target population) are to be extrapolated to the larger population.

Several approaches to population sampling have been identified. Macfarlane et al. (1985) opted to cover a variety of occupations in order that both sedentary and manual workers were included. To achieve this, volunteers from several departments in local government in the region of Strathclyde were sought. The distribution of the 1338 subjects may be seen in Figure 3.1.

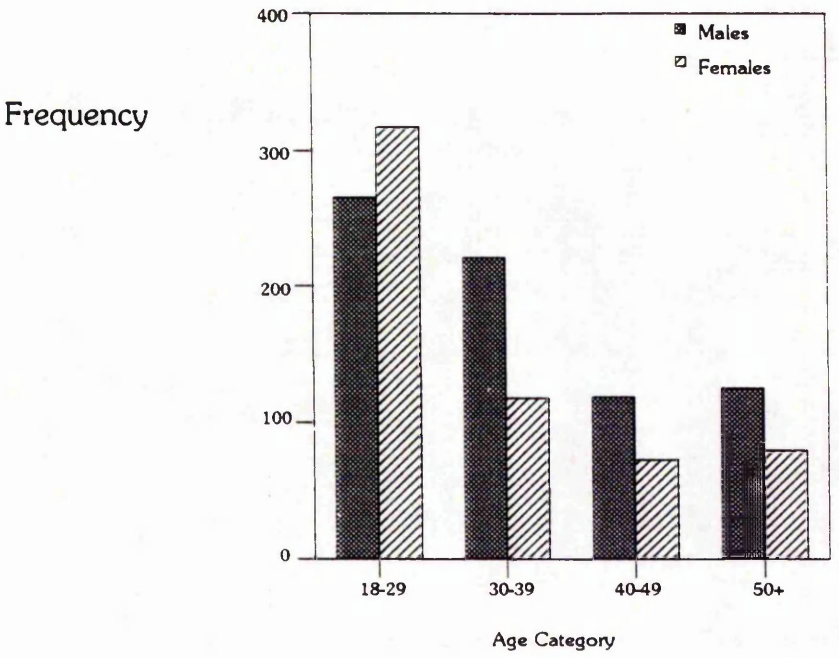


Figure 3.1 **Distribution of 1338 apparently healthy individuals in the Glasgow Study.**

It can be seen that there is a predominance of younger persons in this sample, presumably because they are more willing to volunteer for screening. This can be attributed to the fact that the probability of finding any abnormality of a cardiac nature will be lower in the younger age groups than for older people. It is also clear that there is a large difference in the number of males and females in the 30+ age

categories since many women opt to leave employment at this stage in order to start a family (Macfarlane, 1989d).

Any form of extrapolation from this sample must be questionable since the sample of individuals obtained is not entirely representative of the population in general. In particular, the subgroups specifically thought to be 'at risk', typically the older groups, are not as well represented as the younger subgroups. Therefore any inferences which may be made about such groups should be interpreted with caution although, hopefully, the smoothed age-related upper limits of normality that are produced from this data will not be biased, merely better estimated for the younger ages and for males.

3.3 THE CURRENT SITUATION:

3.3.1 Basic Assumptions.

When considering a large number of observations of a variable of interest on a continuous scale, it is often the case that the resulting frequency distribution demonstrates a spread of values, the majority being in the middle. In the situation where the spread is relatively even about the central value, it is common to assume that the underlying 'population' of the variable will be well approximated by a Normal distribution. This assumption of normality is very important since many statistical techniques are based on this premise. Under this assumption, the normal range, which is commonly taken to be described by the middle 95% of values, can be obtained by calculating the mean and the standard deviation of the sample. Thus the 95% range can be defined as roughly

mean ± 2 standard deviations.

If the underlying distribution of a particular variable is non-Normal, then it may be possible to transform the variable to achieve approximate normality. As a last resort, certain situations may necessitate the use of nonparametric methods of inference to estimate 'normal' or 'healthy' ranges. Such methods do not require strong assumptions as to the underlying distribution which generated the data but are often inferior in providing strong inferences.

It has been established (Simonson, 1961) that the distribution of many ECG variables is skewed so that the normality assumption does not apply. For example, the unconditional distribution of the R wave amplitude in lead V5 is skewed to the right (see Figure 3.2).

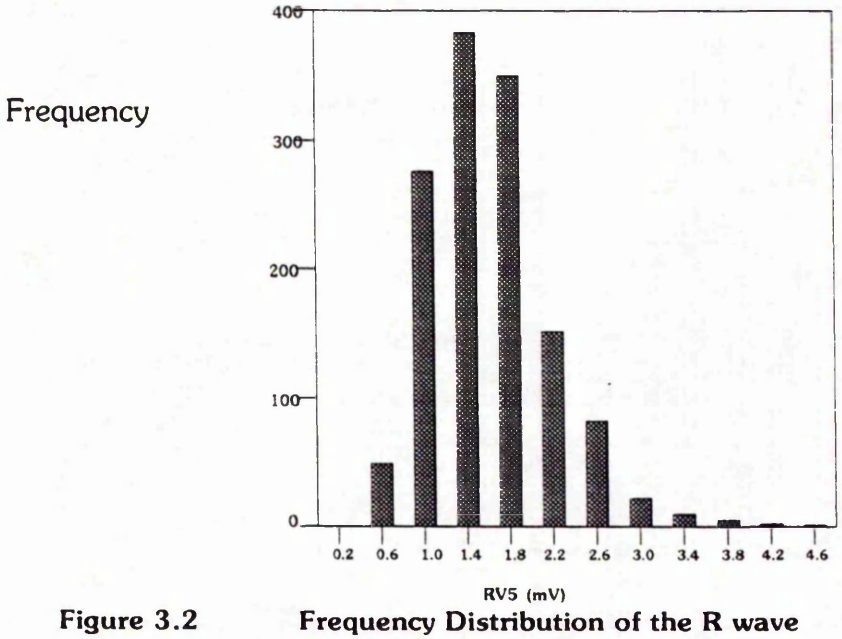


Figure 3.2 **Frequency Distribution of the R wave amplitude in lead V5 (n=1338)**

'Healthy' ranges for the raw age and sex categorised data collected in Glasgow should therefore not be presented as mean ± 2 standard deviations. Instead, 2% of observations were excluded from either end of each group, thereby providing a 96% range. The 96

percentile range is used in many electrocardiographic studies instead of the standard 95% range and there is no special reason for opting to exclude 2% of values rather than 2.5% other than the preservation of continuity.

3.3.2 Discrete Upper Limits.

Upper limits of 'healthiness' for many parameters of interest have been constructed on the basis of the 96% range of the data (Macfarlane, 1989e) and an example of how this limit varies over age and sex for the R wave amplitude in lead V5 can be seen in Figure 3.3.

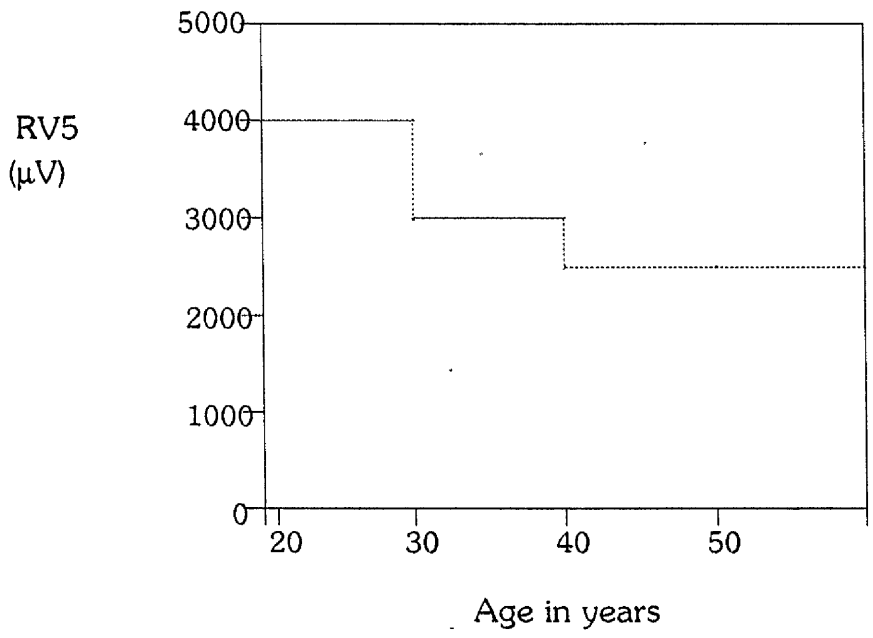


Figure 3.3 **Discrete Upper limits of 'healthiness' for the R wave amplitude in lead V5 (Males)**

As mentioned briefly in the previous chapter, small measurement changes as well as age changes in the neighbourhood of threshold

values are a potential source of lack of repeatability and it is therefore of interest to replace such discrete upper limits of normal with limits which change continuously and smoothly with age.

3.4 SMOOTHING OUT DISCRETE LIMITS:

The use of linear regression.

As an initial step in our main aim of predicting the upper limit of 'healthiness' for a particular ECG variable x, we specify a model to describe the dependence of x on age t. The simplest possible model is a linear regression where the conditional expected value of x depends linearly on t and the variability about such a linear relationship is constant. As in the previous section, the assumption of normality (in this case for the conditional distribution about the expected value) is questionable, so suitable transformations were investigated.

As an example, normality of the conditional distribution of the R wave amplitude in lead V5, given the age, will be assessed.

An initial plot of the amplitude against the age (for males only) indicates that the variance is not constant (see Figure 3.4) since there does seem to be a larger spread of RV5 values for the younger ages.

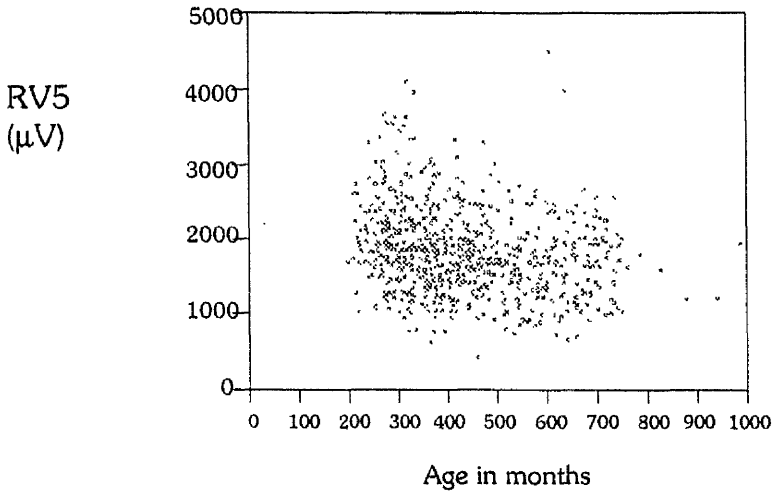


Figure 3.4 **Plot of RV5 Amplitude vs Age in months (Males)**

After applying standard linear regression techniques it is possible to plot the residual values against the fitted values (see Figure 3.5) to assess the adequacy of the linear model.

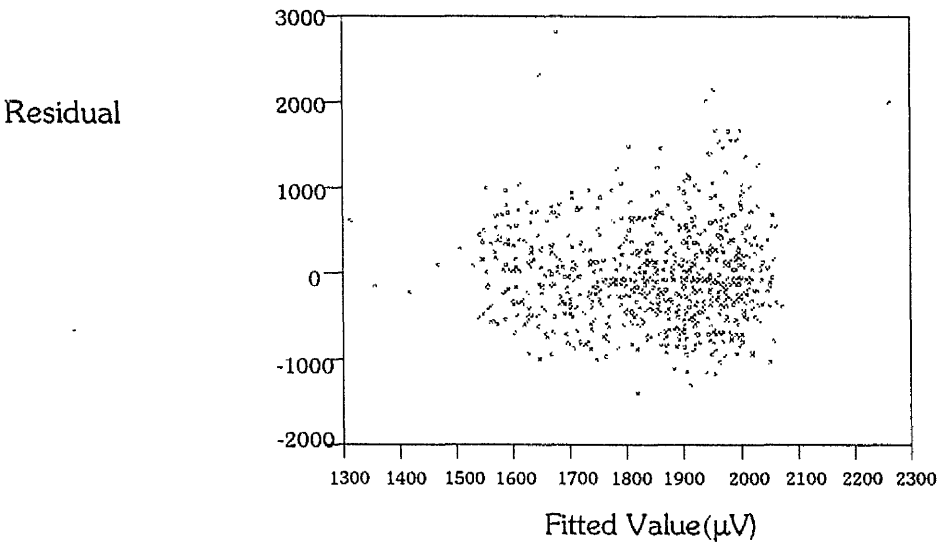


Figure 3.5 **Plot of the fitted value vs. residual for the raw RV5 Amplitude.**

The difference between the true measurement x_i and the fitted value for the i th observation $\hat{x}_i$ defines the i th residual r_i and essentially measures the vertical deviation of measurements from the

fitted line which is defined by the estimated slope $\hat{\beta}$ and intercept $\hat{\alpha}$ (Kalbfleisch, 1979). Thus

$$r_i = x_i - \left(\hat{\alpha} + \hat{\beta} t_i\right).$$

where t_i is the explanatory variable for individual i .

If the model is satisfactory, a plot of the residuals r_i against the fitted values, $\hat{\alpha} + \hat{\beta} t_i$ should demonstrate points lying in a band of constant width about zero exhibiting no patterns or trends. The wedge-shaped nature of Figure 3.5 suggests that there is evidence of increasing variance.

Furthermore, the expected value (ev_i) of the i th observation in a sample of N from a N(0,1) distribution is approximated by

$$ev_i = \Phi^{-1}\{(3i-1)/(3N+1)\}$$

where Φ^{-1} is the inverse Normal cdf.

If our sample had been taken from a Normal distribution, then the plot of observed residuals against these expected values from an N(0,1) distribution would be linear. A normal probability plot (see Figure 3.6) exhibits obvious curvature indicating that the conditional variability is not adequately described by the normality assumption.

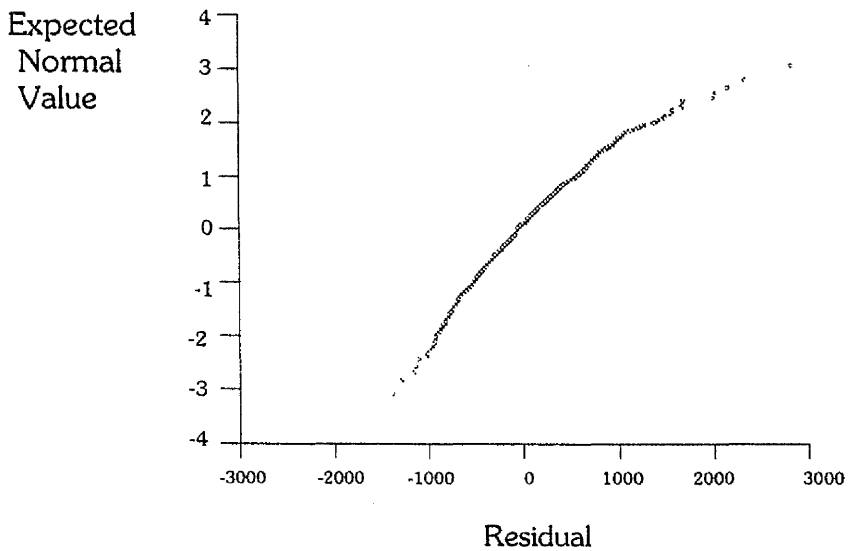


Figure 3.6 **Normal probability plot of the residuals for the raw RV5 Amplitude.**

The evidence of non-constant variance leads us to investigate whether a square root or a logarithmic transformation might be more appropriate to allow us to assume a linear model for the transformed variable.

A plot of the square root of the amplitude of the R wave in lead V5 vs. age for male Caucasians can be seen in Figure 3.7.

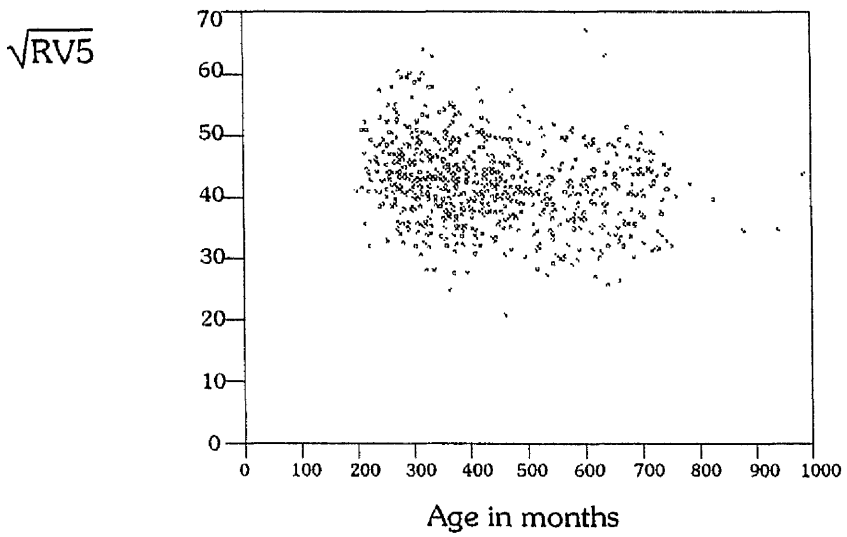


Figure 3.7 **Plot of $\sqrt{RV5}$ Amplitude vs. Age (Males)**

There is possibly a slight suggestion of decreasing variance, but not to the extent that it exists for the raw data.

The residual plot for the transformed data has been provided in Figure 3.8. This demonstrates that the residuals lie in a horizontal band of constant width about zero with no apparent trend. Furthermore, a normal probability plot of the residuals (Figure 3.9) illustrates that the assumption of normality has been satisfied.

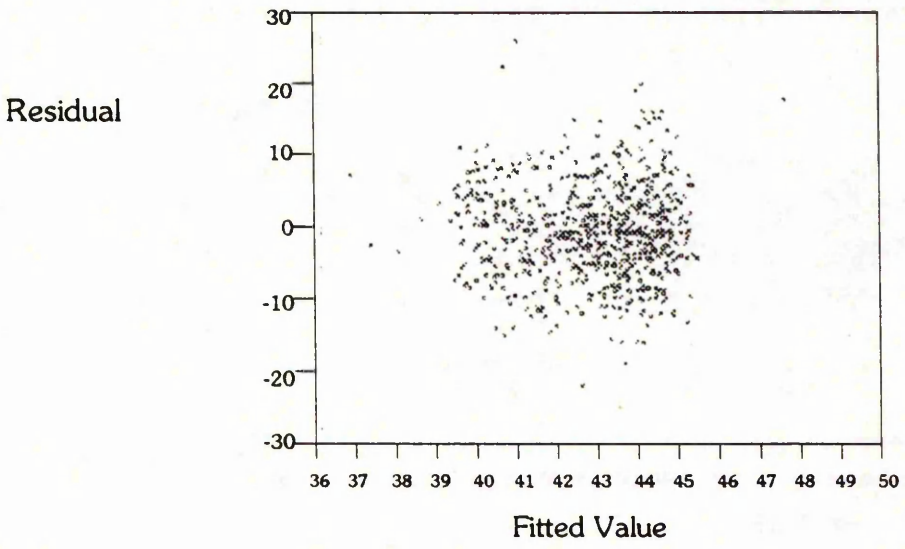


Figure 3.8 **Plot of the fitted value vs. residual for the transformed data (i.e. $\sqrt{RV5}$).**

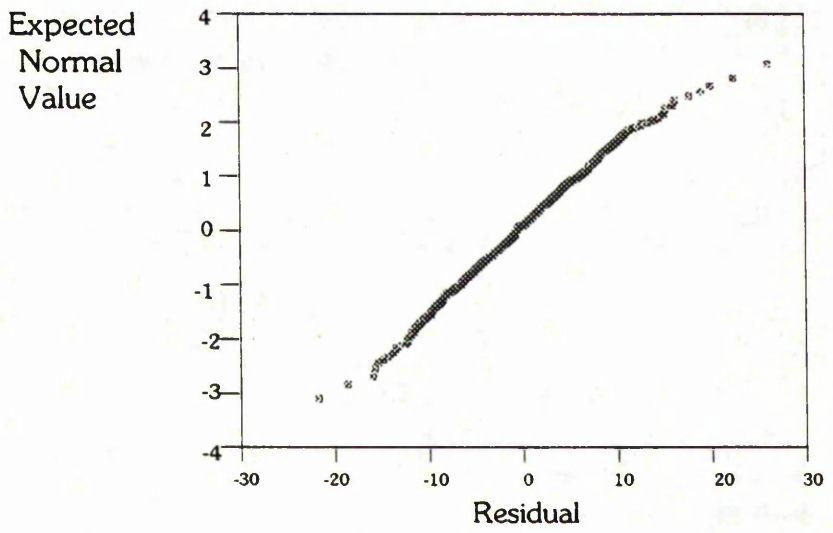


Figure 3.9 **Normal probability plot of the residuals for the transformed data (i.e. $\sqrt{RV5}$).**

Assuming the standard linear model, that $\sqrt{RV5}$ depends linearly on age with constant variance and letting x_i and t_i denote $\sqrt{RV5}$ and age respectively for the i th individual in our basic data set, then

$$E(x_i|t_i) = \alpha + \beta t_i \; ; \qquad Var(x_i|t_i) = \sigma^2$$

where $x_i = \alpha + \beta t_i + \varepsilon_i$, $\varepsilon_i \sim N(0,\sigma^2)$ and x_i, x_j are independent (for $i \neq j$).

The usual least squares estimates were obtained for the unknown parameters α, β, σ and the upper 98 percentile of the distribution (since we are effectively excluding 2% of values from the upper end of the range) can therefore be estimated as:

$$g_{98}(t) = \hat{\alpha} + \hat{\beta}t + t(n-2;0.02)\sqrt{\hat{\sigma}^2\left[1+\frac{1}{n}+\frac{(t-\bar{t})^2}{S_{tt}}\right]} \qquad (3.1)$$

where $S_{tt} = \sum_{i=1}^{n}(t_i - \bar{t})^2$.

Upper limits of 'healthiness' were calculated on the basis of the 1338 ECGs which were recorded from healthy adults from Glasgow (see 3.2).

Equation 3.1 can be approximated by

$$g_{98}'(t) = \hat{\alpha} + \hat{\beta}t + t(n-2;0.02)\hat{\sigma}$$

since for our large samples n and S_{tt} dominate and the terms $\frac{1}{n}$ and $\frac{(t-\bar{t})^2}{S_{tt}} \to 0$.

For the $\sqrt{RV5}$ data, the parameter estimates for males and females are as follows:

	$\hat{\alpha}_j$	$\hat{\beta}_j$	$\hat{\sigma}_j$	$\hat{\alpha} + t(n_j - 2;0.02)\hat{\sigma}_j$
Males	47.25	−0.01089	6.39	59.77
Females	37.40	−0.00273	5.23	47.65

Thus, the equations describing the upper limits of normal for the square root of the R wave amplitude in lead V5 for males and females are approximately

$$\sqrt{RV5} = (59.77 - 0.01089 \times \text{Age})\mu V \quad \text{for males} \qquad (3.2)$$

$$\sqrt{RV5} = (47.65 - 0.00273 \times \text{Age})\mu V \quad \text{for females} \qquad (3.3)$$

Figure 3.10 shows the close similarities between the exact and the approximate upper 98 percentile limits for males.

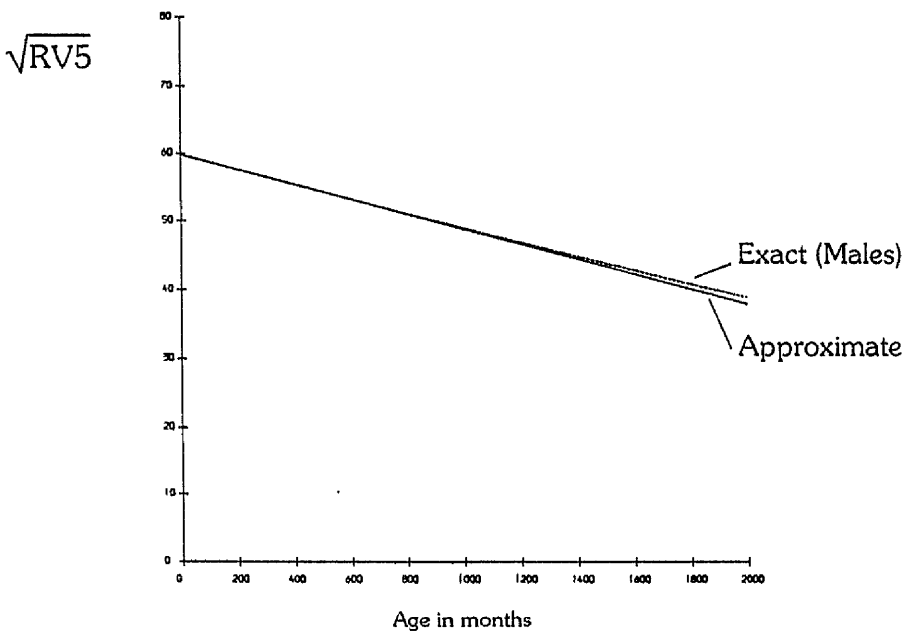


Figure 3.10 Exact and Approximate upper 98 percentile limits for $\sqrt{RV5}$ (Males)

Inferences can now be made about the upper limit of normal for $\sqrt{RV5}$, and hence RV5 simply by squaring the right hand sides of Equations 3.2 and 3.3, for different ages assuming that future

measurements are independent realisations from the same model (i.e. the same population) and that the sample used to produce the equations is representative of a 'healthy' population.

3.5 SMOOTHING OUT DISCRETE LIMITS:

An alternative approach - Nonparametric Regression.

If, even after transformation, we are still not satisfied with the linearity assumption for the conditional expected value in our model then a nonparametric regression model which imposes no linear structure can be fitted. However, the assumptions of conditional normality and non-constant variance in this case are not dropped.

Again, let x_i and t_i denote the ECG variable of interest and age respectively for the i th individual in our data set and assume that inferences about the relationship between x_i and t_i are to be made (for $i = 1,\ldots,n$). We aim to find a smooth function f which represents the upper limit of 'healthiness' for our data. Using a nonparametric approach we can write

$$\mathrm{E}(x_i | t_i) = f(t_i) \qquad \text{and} \qquad \mathrm{Var}(x_i | t_i) = \sigma^2$$

where x_i and x_j are independent (for $i \neq j$) and the only basic requirement of f is that it is 'smooth' (Silverman, 1985).

One immediate and intuitive approach to estimating f is to minimise the sum of squares function:-

i.e.

$$\underset{\text{over all possible } f}{\text{minimise}} \sum_{i=1}^{n} \left(X_i - f(t_i)\right)^2$$

where X_i are observations taken at points t_i $(i = 1,\ldots,n)$.

This expression may, however, achieve the value zero when f is allowed to interpolate the data in the situation where there are no replicated values of t. Clearly this is undesirable and in order to obtain a more realistic estimate of the function f, a 'roughness penalty' incorporating a smoothing parameter δ can be introduced. The smoothing parameter essentially determines the width of the window covering the observations which are currently being used to determine the smoothed curve at any particular choice of t. The more observations contained in the window, the smoother the plot.

The sum of squares function is penalised in the following way:

$$\sum_{i=1}^{n}\left(X_i - f(t_i)\right)^2 + \delta\int_a^b f''(t)^2\,dt$$

where δ dictates the amount of smoothing to be used, $\int_a^b f''^2$ is a convenient form of 'roughness penalty' (Silverman, 1986) and $[a,b]$ is the range of the explanatory variable t.

Effectively, as $\delta \to 0$, f interpolates the data and as $\delta \to \infty$, $f(t) \to \bar{x}$.

For our particular set of data where x is the square root of the R amplitude in lead V5 and t is the age, Figure 3.11 shows upper 98-percentile curves which have been based on nonparametric regression methods and have been produced for two different values of δ.

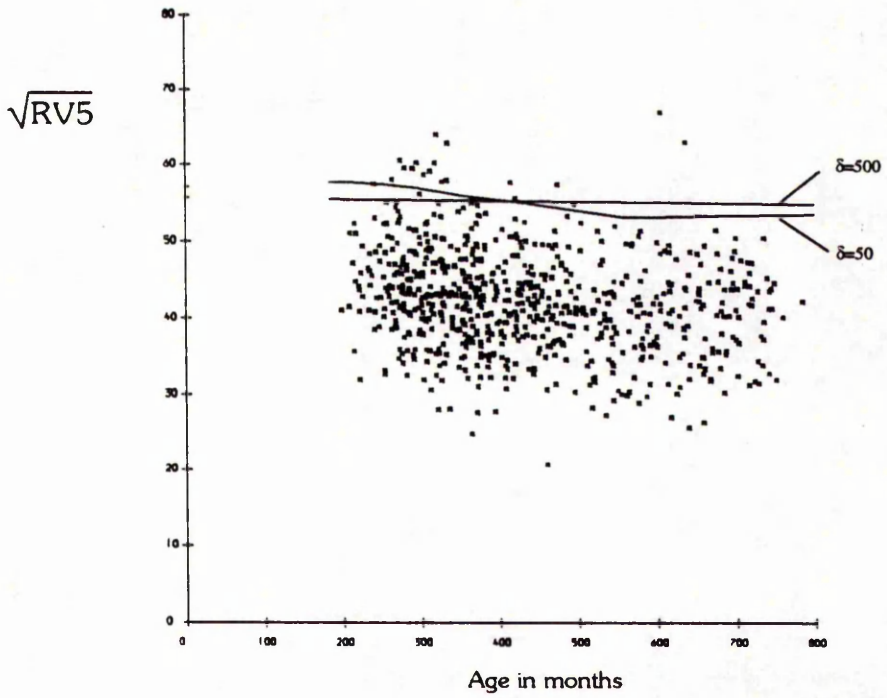


Figure 3.11 **Upper 98-percentile Curves for δ=50 and δ=500 based on Nonparametric Regression.**

The larger value of δ produces the smoother curve. Cross-validation is a popular technique for choosing the 'best' value of δ and works on the principle that each data point is omitted in turn and predicted on the basis of the remaining data. The optimal value of δ is that value which predicts the remaining data points most accurately.

The 'smooth' function will be estimated as follows:

$$\hat{f}(t) = \sum_{i=1}^{n} x_i w_i(t)$$

where $\sum_{i=1}^{n} w_i(t) = 1$, $w_i(t) \geq 0$ and $w_i(t) \propto \frac{1}{\delta} K\left(\frac{t - t_i}{\delta}\right)$,

i.e. a 'simple' weighted average of the $\sqrt{RV5}$ values where the weights depend on the age under consideration.

A variety of techniques for producing nonparametric estimates is available ranging through splines, kernels, etc. We have chosen a

kernel method because of its simplicity and hence its easy application.

Here K denotes a kernel function which is symmetric and centred on the origin, satisfying the condition $\int_{-\infty}^{\infty} K(z)\mathrm{d}z = 1$. A Gaussian (or Normal) kernel has been used to construct an estimate for $f(t)$. Thus we choose

$$f(t) = \frac{\sum_{i=1}^{n} x_i e^{\frac{-1}{2}\frac{(t-t_i)^2}{\delta^2}}}{\sum_{i=1}^{n} e^{\frac{-1}{2}\frac{(t-t_i)^2}{\delta^2}}}$$

as a suitable nonparametric regression of x on t with smoothing parameter δ.

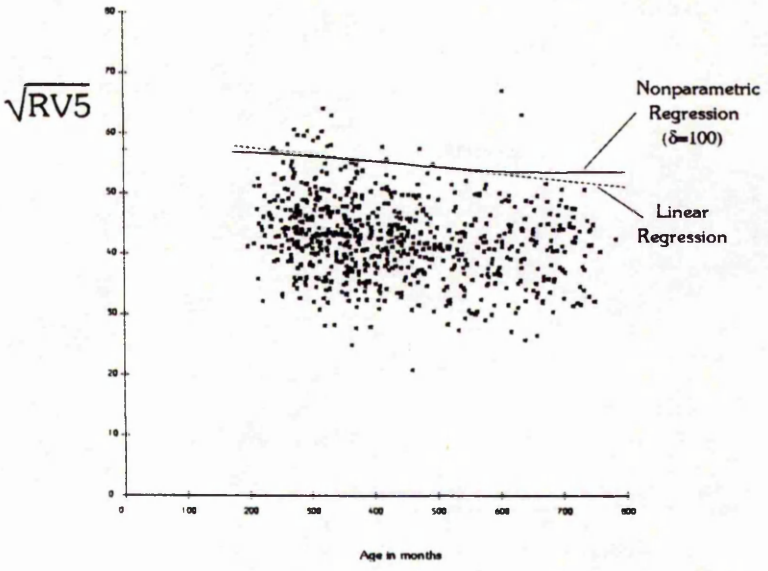


Figure 3.12 **A comparison of the upper limits of 'healthiness' calculated using a nonparametric regression technique and a simple linear regression technique.**

Figure 3.12 demonstrates the similarity between the nonparametric curve and the linear regression curve obtained for the square root of the R wave amplitude in lead V5 (for male

Caucasians), clearly suggesting that the parametric linear regression previously described is effective over most of the age range and is certainly a simpler alternative. This also applies to many other ECG measurements (see examples in Figures 3.13 to 3.15). Thus the nonparametric approach has been a beneficial exploratory exercise.

3.6 THE EFFECT OF SMOOTHING IN PRACTICE.

Figures 3.13 to 3.15 demonstrate how the continuous equations describing the upper limit of 'healthiness' of three voltage measurements compare with the discrete limits. The points of discontinuity which exist for the discrete limits can give rise to a lack of repeatability if, for example, a patient's age group changes from one recording to the next. The use of continuous equations eliminates this discontinuity by providing a method of describing the behaviour of the measurement of interest (or a transformation thereof) smoothly over the age range.

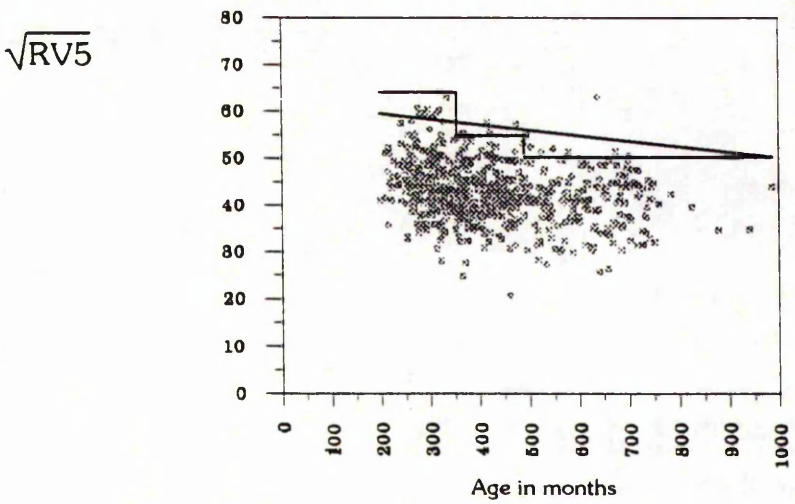


Figure 3.13 **The smooth upper limit of 'healthiness' for the square root of the R wave amplitude in lead V5 compared to the original discrete upper limit (Males).**

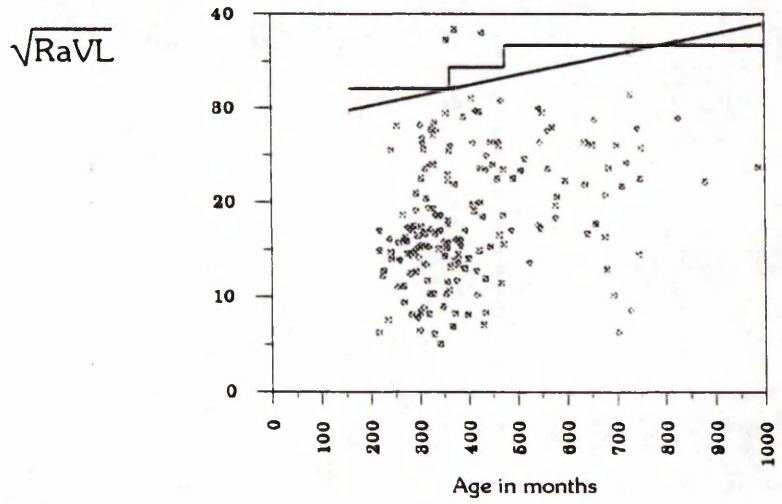


Figure 3.14 **The smooth upper limit of 'healthiness' for the square root of the R wave amplitude in lead aVL compared to the original discrete upper limit (Males).**

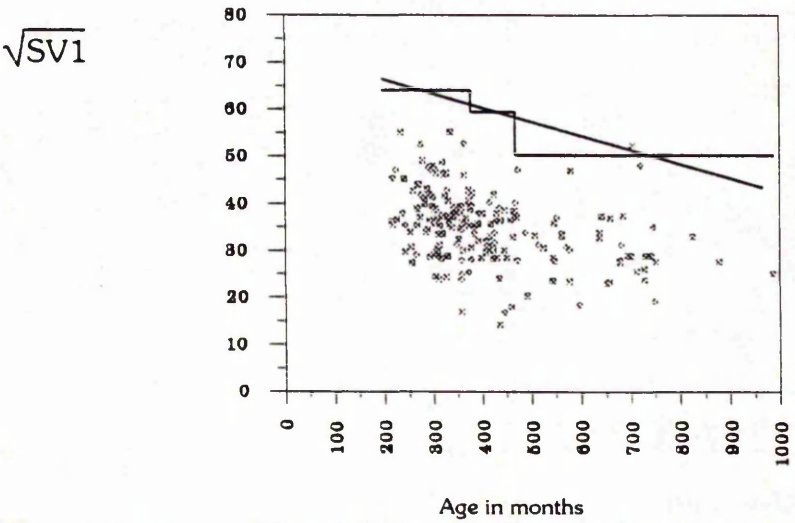


Figure 3.15 **The smooth upper limit of 'healthiness' for the square root of the S wave amplitude in lead V1 compared to the original discrete upper limit (Males).**

Equally, these newly established continuous upper limits of 'healthiness' may be applied to other ECG measurements. In general, limits for wave amplitudes are based on the square root of the measurement whereas limits for durations can be constructed using the raw data without first transforming. Figure 3.16 illustrates that

the unconditional distribution of the QRS duration in lead V5 is approximately Normal.

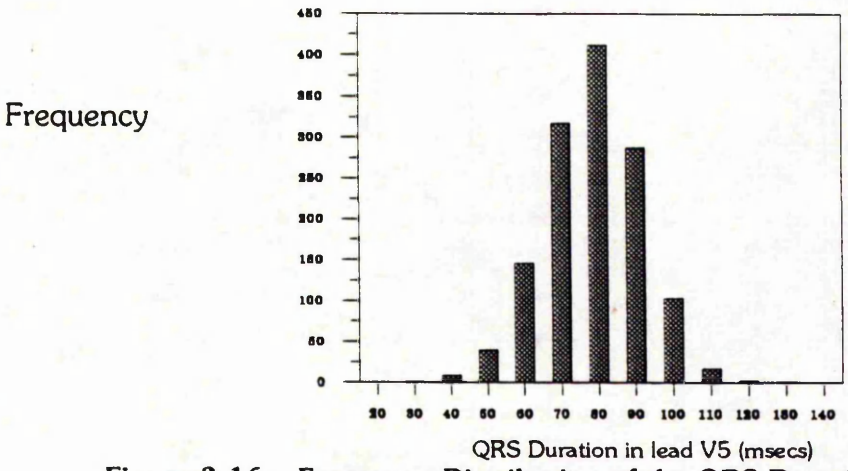


Figure 3.16 **Frequency Distribution of the QRS Duration in lead V5**

Figure 3.17 demonstrates that the variation in the QRS duration in lead V5 is reasonably constant throughout the age range and Figures 3.18 and 3.19 further verify the assumption of normality.

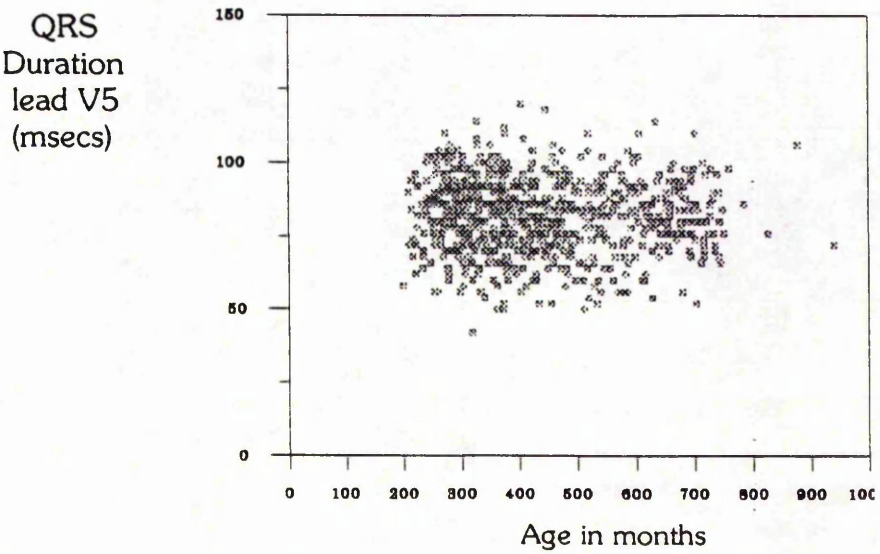


Figure 3.17 **Plot of QRS Duration in lead V5 vs. Age in months (Males)**

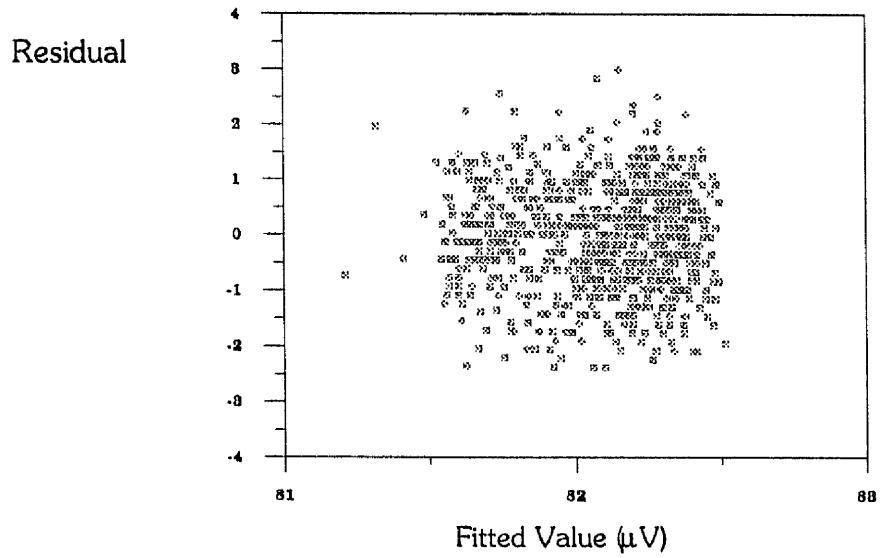


Figure 3.18 **Plot of the fitted value vs. residual for QRS Duration in lead V5**

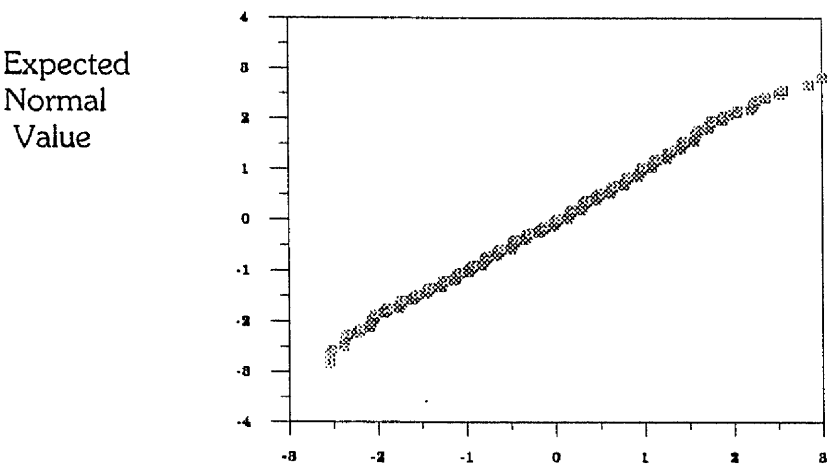


Figure 3.19 **Normal probability plot of the residuals for QRS Duration in lead V5**

3.7 SUMMARY.

For many ECG variables, notably the amplitudes of the waveforms, it was found that by taking square roots a simple linear

model to relate the transformed ECG variable to age was adequate. In the case of wave durations, no transformations were required. Thus continuous upper limits of 'healthiness' using age as the explanatory variable were easily achieved. These limits can help to minimise the lack of consistency between recordings due to their smooth behaviour throughout the age range and help to form the basis of a more repeatable approach to ECG interpretation.

CHAPTER FOUR:

SOURCES OF REPEAT VARIABILITY IN THE ECG:

Day-to-day variation.

4.1 INTRODUCTION.

Lack of consistency between diagnoses based on successive ECG recordings can arise as a result of many sources of variation. Day-to-day variation which has been mentioned briefly in Chapter 2 can result in conflicting diagnoses if small ECG measurement changes cause the thresholds which distinguish 'healthy' from 'unhealthy' to be crossed between consecutive recordings. This repeat variation may be the direct result of many contributory factors which have been described previously. However, it would be unrealistic and impractical to attempt to collect data on all such sources of variability since a model incorporating separate estimates of variation due to respiratory stage, the effect of smoking, level of anxiety of the patient etc. would be too complicated to be of any practical use. A much simpler option is to concentrate on estimating overall day-to-day variation which will include factors such as electrode positioning, recording technique and anxiety level of the patient from one day to the next in a single estimate.

Here, we will consider how to model day-to-day variation and describe methods for the estimation of the amount of variability which exists from one recording to the next in our particular set of

data comprising replicate ECGs which were recorded at least 24 hours apart.

4.2 THE DATA USED TO ESTIMATE DAY-TO-DAY VARIATION.

To obtain estimates of the day-to-day variability (σ_x) for a particular measurement of an ECG variable x, an adequately sized database of ECG recordings representing a variety of cardiac pathologies is required. Furthermore, the patients must be regarded as being 'ECG-stable', i.e. recordings are to be taken from such individuals whose clinical condition is such that there is no medical reason for their ECGs to change from one day to the next.

ECGs were recorded at least 24 hours apart from non-acute cardiac patients admitted to Glasgow Royal Infirmary between August 1988 and December 1991. To represent real day-to-day variation, no standardising procedures were used, i.e. electrode positions were not marked nor were there any conditions placed on the recording technicians (i.e. the technician recording the ECG on day 2 need not necessarily be the same as the one who recorded the initial ECG). In this way, an estimate of day-to-day variability included components of variability due to recording technique, electrode positioning, respiratory phase, posture and anxiety level.

	No.	Mean Age	Min. Age	Max. Age
Male	234	56	23	80
Female	61	58	29	76

Table 4.1 **Mean, maximum and minimum ages of the 295 patients used to construct estimates of day-to-day variation**

There is a predominance of males in this database as can be clearly seen in Table 4.1. This is attributable to the fact that it is mainly men who are admitted to the University Department of Medical Cardiology at Glasgow Royal Infirmary suffering from non-acute cardiac conditions, predominantly ischæmic heart disease.

4.3 INITIAL INVESTIGATION OF THE DAY-TO-DAY VARIABILITY.

It is unlikely that an observed value of a day 1 ECG measurement will always be identical to the observed value of the day 2 reading of the same variable. We would like to be able to estimate the amount of variation between such readings in a normal population. It is reasonable to assume that the day-to-day variation may depend on the magnitude of the measurement being considered with large values being associated with greater day-to-day variability than small values.

To investigate how the day-to-day variability behaved in terms of any dependence on the magnitude of a particular ECG variable, some simple plots were considered. The initial step was to take differences of the day 1 and day 2 readings in order to remove any patient effect leaving the differences reflecting only measurement error. Plotting the day 1 measurement of a particular variable against the difference between day 1 and day 2 provided an initial impression of whether there was any relationship between variability and magnitude present. In some cases, the difference in the measurements between recordings was constant over the entire range of the day 1 readings (see Figure 4.1).

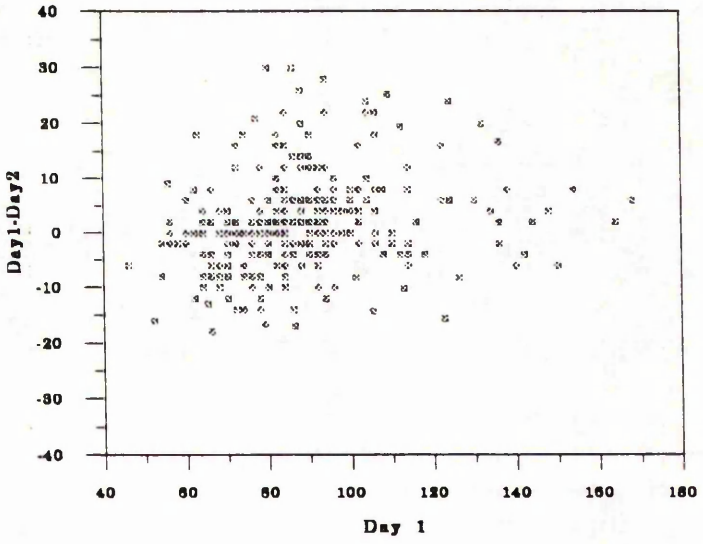


Figure 4.1 **QRS V5 Duration (msecs)**
Day1 vs. difference between Day1 and Day2

In this scenario, a possible model for the differences d between day 1 and day 2 measurements x_1 and x_2 respectively for patient i could be constructed as follows:

$$\begin{aligned} &\text{DAY 1} && x_1 = \mu + \varepsilon_1 \\ &\text{DAY 2} && x_2 = \mu + \varepsilon_2 \end{aligned}$$

where μ is the average 'patient i ' value ($i = 1,\ldots,n$) and ε_1 and ε_2 are independent daily measurement errors ($\varepsilon_j \sim \mathrm{N}(0,\sigma_x^2)$).

The difference between the day 1 and day 2 measurements may be calculated as:

$$\begin{aligned} d &= x_1 - x_2 \\ &= (\mu + \varepsilon_1) - (\mu + \varepsilon_2) \\ &= \varepsilon_1 - \varepsilon_2 \end{aligned}$$

where $d \sim \mathrm{N}(0, 2\sigma_x^2)$.

However, certain other variables demonstrated a trend whereby the larger the day 1 measurement, or indeed the magnitude of the ECG variables (as measured by the average of the two readings), the greater the likely variability between measurements (see Figure 4.2).

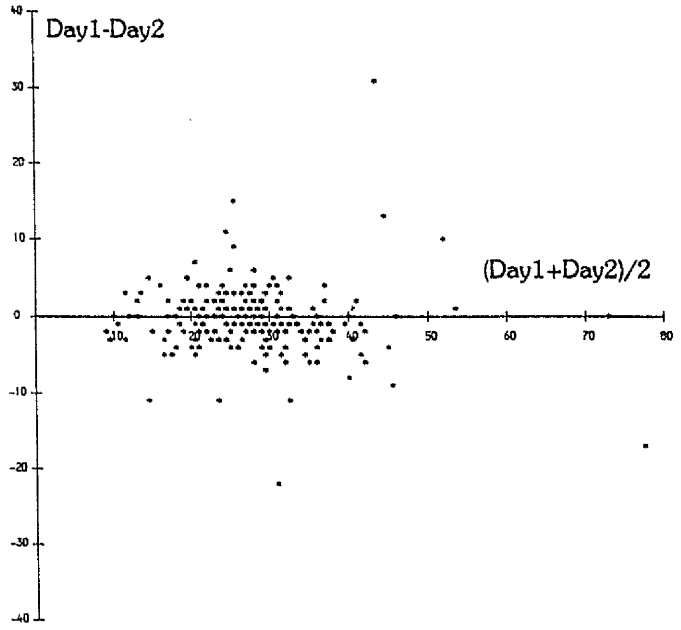


Figure 4.2 **RI Duration (msecs)**
(Day1+Day2)/2 vs. Day1-Day2

It is more appropriate to plot the average of the two readings against the difference, since there may be a spurious correlation between the day 1 value and the difference.

In this situation, a possible model for the difference d between day 1 and day 2 observations x_1 and x_2 respectively is of the form:

$$\begin{array}{ll} \text{DAY 1} & x_1 = \mu + \varepsilon_1 \\ \text{DAY 2} & x_2 = \mu + \varepsilon_2 \end{array}$$

where ε_1 and $\varepsilon_2 \sim N(0, \sigma_x^2 + \tau_x^2 \mu)$, $d \sim N(0, 2(\sigma_x^2 + \tau_x^2 \mu))$, and μ is the mean of the day 1 and the day 2 readings.

It is therefore reasonable to assume that there may be some form of relationship present between σ_x and the ECG variable of interest (x). For this reason, the relationship between variability and magnitude must be considered when attempting to estimate σ_x.

4.4 A PROPOSED MODEL FOR DAY-TO-DAY VARIABILITY

4.4.1 The Basic Model.

When modelling the observations in our data set, it is often appropriate to assume that the underlying distribution of the measurements of interest is approximately Normal. Many sets of experimental data exhibit properties of a random sample which has been taken from a Normal distribution, often after a suitable transformation.

Suppose we have data available for the ith patient $(i=1,\ldots,n)$ with reading x_{i1} on day 1 and reading x_{i2} on day 2. Since day-to-day variability appears to be related to the magnitude of the variable in certain cases, we propose the following model:

$$\begin{array}{ll} \text{DAY 1} & x_{i1} = \mu_i + \varepsilon_{i1} \\ \text{DAY 2} & x_{i2} = \mu_i + \varepsilon_{i2} \end{array}$$

where $\varepsilon_{ij} \sim \mathrm{N}\left(0, \left(\sigma^2 + \tau^2 \mu_i\right)\right)$ for $i=1,\ldots,n$; $j=1,2$; and μ_i is the true value of patient i at that time. Here, σ^2 measures the baseline variability while τ^2 measures the rate of increase of the variance with the magnitude of the ECG variable x. This model will apply to both of the situations which have previously been mentioned (i.e. when the difference between the day 1 and day 2 measurements depends on the day 1 value or not) simply by specifying $\tau=0$ or $\tau>0$ as appropriate.

(Note that the proposed model is of the form $\sigma_x^2 = \sigma^2 + \tau^2 \mu_x$ although other models, such as $\sigma_x = \sigma + \tau \mu_x$, might be equally sensible).

All of the μ_i (i.e. the true values of each of the n patients) are nuisance parameters since interest lies in estimates of σ and τ, not in the particular patients. If in fact $\tau = 0$ then a simple way of removing the dependence of the model on these superfluous parameters is to take differences, i.e.

$$\begin{aligned} d_i &= x_{i1} - x_{i2} \\ &= \varepsilon_{i1} - \varepsilon_{i2} \end{aligned}$$

and $$d_i \sim \mathrm{N}\left(0, 2\sigma^2\right) \qquad (\text{for } \tau = 0).$$

So if we extend this to the case of $\tau \neq 0$, taking differences will remove part of the dependence on the μ_i,

i.e.

$$\begin{aligned} d_i &= x_{i1} - x_{i2} \\ &= \varepsilon_{i1} - \varepsilon_{i2} \end{aligned}$$

and $$d_i \sim \mathrm{N}\left(0, 2\left(\sigma^2 + \tau^2 \mu_i\right)\right)$$

Basing our inferences on the differences, a convenient log-likelihood function can be written as

$$l\left(\sigma, \tau, \mu_i; \underline{x}\right) = -\frac{1}{2} \sum_{i=1}^{n} \log\left(\sigma^2 + \tau^2 \mu_i\right) - \sum_{i=1}^{n} \frac{\left(x_{i1} - \mu_i\right)^2 + \left(x_{i2} - \mu_i\right)^2}{2\left(\sigma^2 + \tau^2 \mu_i\right)} \qquad (4.1)$$

and the parameter values $(\hat{\sigma}, \hat{\tau})$ which maximise Equation 4.1 are the joint maximum likelihood estimates (MLEs) of σ and τ. These estimates together will correspond to the most plausible values of (σ, τ).

One approach to eliminating the remaining nuisance parameters μ_i, is to maximise the above log-likelihood function $l(\sigma,\tau,\mu_i;\underline{x})$ over μ_i to yield a function of σ,τ only. The values of μ_i as functions of σ and τ which maximise the above expression will provide the Profile Likelihood for σ, τ (Kalbfleisch, 1979) which can then be used for inferences on σ and τ and hence maximum (profile) likelihood estimates $(\hat{\sigma},\hat{\tau})$, i.e.

$$\begin{aligned} pl\,(\sigma,\tau;\underline{x}) &= \max_{\mu_i}\; l\,(\sigma,\tau,\mu_i;\underline{x}) \\ &= \quad l(\sigma,\tau,\hat{\mu}_i(\sigma,\tau);\underline{x})\ . \end{aligned}$$

However, trying to achieve this by maximising Equation 4.1 for fixed σ and τ, we find that the appropriate equation $\frac{\partial l}{\partial \mu_i}=0$ cannot be solved analytically. We would therefore have to use a numerical method of evaluating each $\hat{\mu}_i(\sigma,\tau)$ for every pair of values of σ and τ and this would be computationally extensive.

To avoid this, one appealing approach is to obtain an approximation to the Profile Likelihood by estimating $\hat{\mu}_i(\sigma,\tau)$ from the data simply as the sample average, i.e. by $m_i=\frac{(x_{i1}+x_{i2})}{2}$.

Thus by replacing μ_i by m_i in Equation 4.1 above, the approximate profile likelihood can be written:

$$\hat{pl}(\sigma,\tau;\underline{x})=l(\sigma,\tau,m_i;\underline{x}) \qquad \text{where } m_i=\frac{(x_{i1}+x_{i2})}{2},$$

i.e.

$$\hat{pl}(\sigma,\tau;\underline{x})=-\frac{1}{2}\sum_{i=1}^{n}\log(\sigma^2+\tau^2 m_i)-\sum_{i=1}^{n}\frac{d_i^{\,2}}{4(\sigma^2+\tau^2 m_i)}$$

where $d_i=(x_{i1}-x_{i2})$.

If indeed this form of profile likelihood is a reasonable approximation we can then proceed to base our inferences on σ and τ on this computationally much simpler function.

4.4.2 Comparing the Approximate and Exact profile log-likelihoods.

Before proceeding, we first consider how well this approximation to the profile log-likelihood compares to the exact method,

i.e. compare

$$\hat{pl}(\sigma,\tau;\underline{x}) = l(\sigma,\tau,m_i;\underline{x})$$

with

$$pl(\sigma,\tau;\underline{x}) = l(\sigma,\tau,\hat{\mu}_i(\sigma,\tau);\underline{x})$$

where all the μ_i (i=1,..., n) are maximised at each of all the possible pairs of values of (σ,τ).

The performance of the approximation can be assessed by plotting the exact Profile Likelihood (which requires considerable computation) over a plausible range of σ, τ and examining the discrepancies between it and the approximate Profile Likelihood derived by replacing $\hat{\mu}_i(\sigma,\tau)$ by m_i.

If the maximum percentage difference between the exact log likelihood and the approximate log likelihood over a sensible range of (σ,τ) is less than about 1% then it seems reasonable to assume that the approximation is adequate,

i.e. if $$\max_{(\sigma,\tau)} \frac{pl(\sigma,\tau;\underline{x}) - \hat{pl}(\sigma,\tau;\underline{x})}{pl(\sigma,\tau;\underline{x})} < 0.01$$

then the approximation is assumed adequate.

4.4.3 Illustration 1 :

The R wave duration in lead V1.

The approximate log-likelihood function has been calculated for the R wave duration in lead V1. The surface plot which has been constructed over a suitable range of (σ,τ) can be seen in Figure 4.3. In addition, curves of $\hat{pl}(\sigma,\tau;\underline{x})=h$ for certain values of h may be provided, thereby producing a contour map of the approximate log-likelihood. The MLE together with approximate 99% and 95% confidence intervals for the approximate log-likelihood for the RV1 duration can be seen in Figure 4.4.

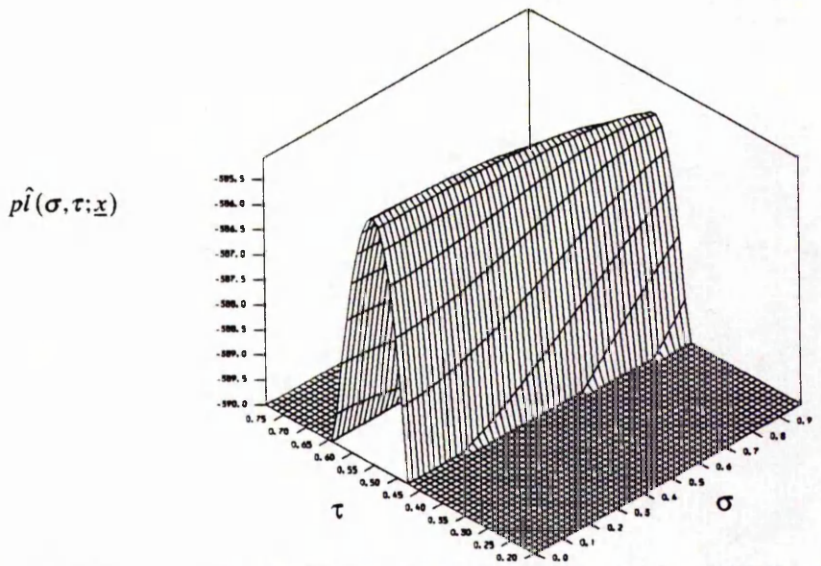


Figure 4.3 **Surface plot of the Approximate log-likelihood function for the RV1 duration.**

(Note that log-likelihood values of less than -390 have been set constant at this value in order to give a clearer picture of the values of σ,τ surrounding the maximised log-likelihood).

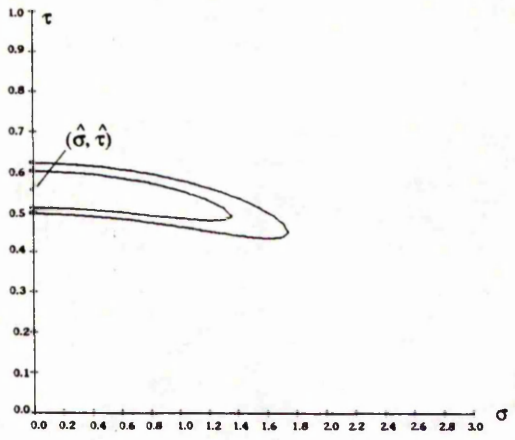


Figure 4.4 **Approximate 99% and 95% Confidence Regions for (σ,τ) (RV1 duration)**

At the 5% significance level we may reject the hypothesis that $\tau = 0$ since the value lies outside the approximate 95% Confidence region in Figure 4.4.

The approximate log-likelihood function for the R wave duration in lead V1 is maximised at

$$(\hat{\sigma}, \hat{\tau}) = (0.01, 0.56).$$

Using the exact log-likelihood function, the MLE is (0.02,0.56). The resulting surface and contour plots can be seen in Figures 4.5 and 4.6.

$pl(\sigma, \tau; \underline{x})$

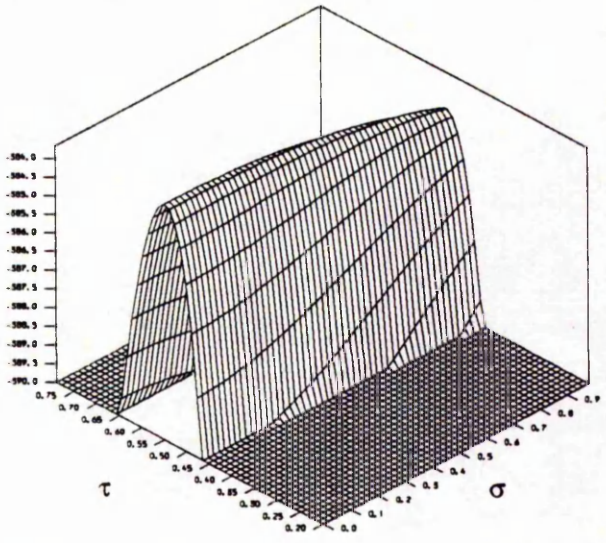


Figure 4.5 **Surface plot of the Exact log-likelihood function for the RV1 duration.**

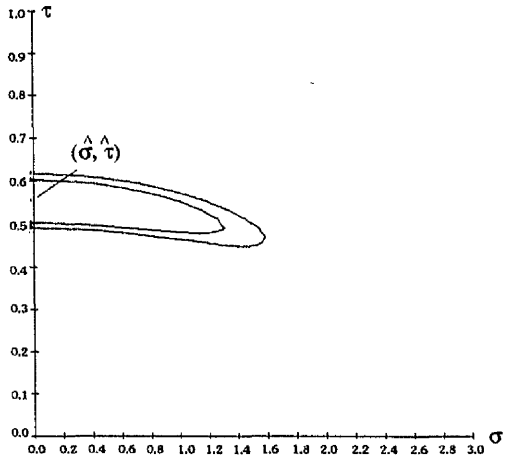


Figure 4.6 **Exact 99% and 95% Confidence Regions for (σ,τ) (RV1 duration)**

From the diagrams there appears to be very little difference between the approximate and the exact maximum likelihood estimates of (σ,τ) for the R wave duration in lead V1. Since the maximum percentage difference between the exact and the approximate log-likelihood values is around 0.41% it is therefore reasonable to use the approximation as an alternative without any significant loss of accuracy.

4.4.4 Illustration 2 :

The R wave amplitude in lead V5.

The approximate log-likelihood function has been calculated for the R wave amplitude in lead V5. The surface plot which has been constructed over a suitable range of (σ,τ) is shown in Figure 4.7. The MLE together with 99% and 95% confidence intervals for the approximate log-likelihood for RV5 can be seen in Figure 4.8.

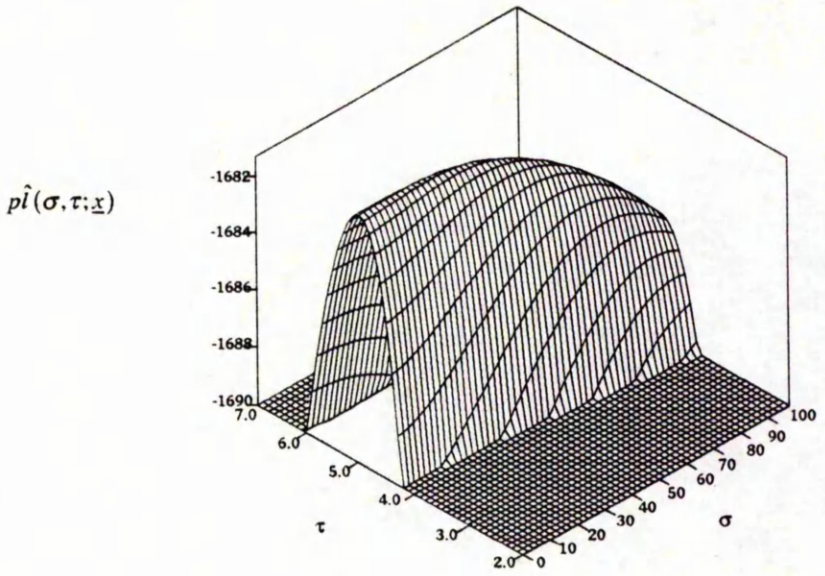


Figure 4.7 **Surface plot of the Approximate log-likelihood function for RV5 Amplitude**

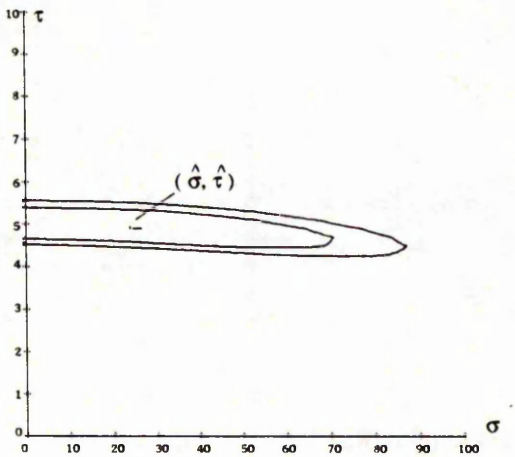


Figure 4.8 **Approximate 99% and 95% Confidence Regions for (σ,τ) (RV5 Amplitude)**

There is evidence that $\tau > 0$ since the value $\tau = 0$ lies well outside the approximate 95% confidence interval in Figure 4.8. Although we cannot conclusively reject at the 5% significance level that $\sigma = 0$ since the 95% contour intersects the line $\sigma = 0$, there is still evidence that $\sigma > 0$ since the value $\sigma = 0$ only just lies inside the approximate 95% confidence interval in Figure 4.8.

The approximate log-likelihood function for RV5 is maximised at $(\hat{\sigma},\hat{\tau})=(28.0,5.0)$, whereas the exact log-likelihood is maximised at $(\hat{\sigma},\hat{\tau})=(21.0,4.8)$. The maximum percentage difference between the log-likelihoods calculated by the two methods is 0.7%, again proving that the approximate method is adequate.

4.4.5 When there is a trend from Day 1 to Day 2: An Alternative Model.

In general, for most ECG variables, the model defined in 4.4.1 proved adequate when applied to the database with either $\tau=0$ or $\tau>0$ as appropriate. However, perhaps contrary to our expectations, we detected that for several variables there was the possibility of a trend from the day 1 to day 2 recordings. For example, the difference observed between day 1 and day 2 measurements from the ST-T segment and from the P wave did not appear to depend on the initial value in the way described in 4.4.1 (i.e. with the expected value of the difference equal to 0 consistently across the magnitude of the measurement). Instead, 'high values' on day 1 tended to produce 'low values' on day 2 and vice versa.

A plot of the differences between recordings versus the day 1 value of the duration of the negative P wave in lead V1 demonstrates a trend which may be interpreted as a 'regression to the mean' (see Figure 4.9). However, we do not observe this phenomenon when we plot the average of the day 1 and day 2 values against their difference. This provides evidence that a spurious correlation does exist for this parameter.

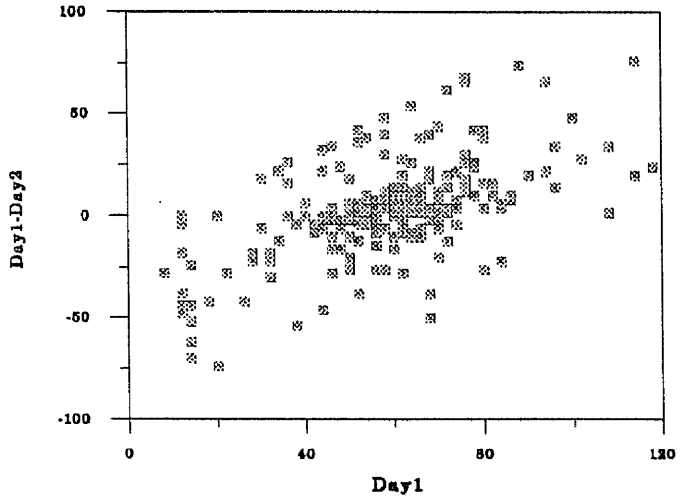


Figure 4.9 **Negative P wave duration (msecs) in lead V1 Day 1 vs. Difference between Day 1 and Day 2.**

However, since the 'regression to the mean' phenomenon appears to happen in practice when considering the ST segment, we must allow for such a dependence in estimating the day-to-day variability and so an alternative model must be investigated. Again, assuming that the underlying distribution of the measurement of interest is approximately Normal, the following may be considered an appropriate model where the mean of each day 2 observation is reduced if the day 1 reading for that patient is above the population average (i.e. $\overline{\mu}$) and increased if the day 1 reading is below average.

$$\text{DAY1} \qquad x_{i1} = \mu_i + \varepsilon_{i1}$$

$$\text{DAY2} \qquad x_{i2} = \mu_i - \lambda(\mu_i - \overline{\mu}) + \varepsilon_{i2}$$

where $\varepsilon_{ij} \sim \text{N}(0, (\sigma^2 + \tau^2 \mu_i))$ for $i = 1,\ldots,n;\ j = 1,2$ and $\lambda > 0$ is a measure of the rate of 'regression to the mean'. If $\lambda = 0$, then the phenomenon does not exist.

Nuisance parameters again cause problems, complicated by the fact that there is the extra nuisance parameter λ. Again, the exact profile likelihood for (σ, τ) could be evaluated numerically but would be even more computationally intensive because of the additional

presence of λ which also cannot be estimated in analytic form. Therefore we proceed to derive an approximation that can be dealt with in a reasonably simple manner as before.

Taking differences as before eliminates these nuisance parameters to a certain extent so that

$$\begin{aligned} d_i &= x_{i1} - x_{i2} \\ &= \lambda(\mu_i - \overline{\mu}) + (\varepsilon_{i1} - \varepsilon_{i2}) \end{aligned}$$

and

$$d_i \sim \mathrm{N}\left(\lambda(\mu_i - \overline{\mu}), 2(\sigma^2 + \tau^2\mu_i)\right)$$

Further manipulation is required to eliminate the presence of λ and the μ_i in the distribution of the d_i. The first step in achieving this is to subtract an amount $\lambda(\mu_i - \overline{\mu})$ from d_i above since, if indeed the parameters were known, this would eliminate the nuisance parameters from the means of the d_i.

So we will construct d_i^* where

$$\begin{aligned} d_i^* &= d_i - \lambda(\mu_i - \overline{\mu}) \\ &= \lambda(\mu_i - \overline{\mu}) + (\varepsilon_{i1} - \varepsilon_{i2}) - \lambda(\mu_i - \overline{\mu}) \\ &= \varepsilon_{i1} - \varepsilon_{i2} \\ &\sim \mathrm{N}\left(0, 2(\sigma^2 + \tau^2\mu_i)\right) \end{aligned}$$

However, in practice we must find some form of estimators for λ and μ_i and hence construct $\hat{d}_i^*$ such that

$$\hat{d}_i^* = d_i - \hat{\lambda}\left(\hat{\mu}_i - \hat{\overline{\mu}}\right),$$

i.e. replacing $\lambda, \mu_i, \overline{\mu}$ by appropriate estimators in the definition above.

The parameter λ will be estimated by fitting a regression of the d_i on the $\left(x_{i1} - \overline{\hat{\mu}_{D1}}\right)$ (i.e. without any intercept) where x_{i1} are the day 1 measurements for each patient and $\overline{\hat{\mu}_{D1}}$ is the average of all the day 1 measurements. A plot of these transformed differences vs. the day 1 measurement of the duration of the negative component of the P wave in lead V1 is provided in Figure 4.10. This is merely for illustrative purposes, since the methods described will be applied to the ST segment only, and not to the P wave duration in lead V1.

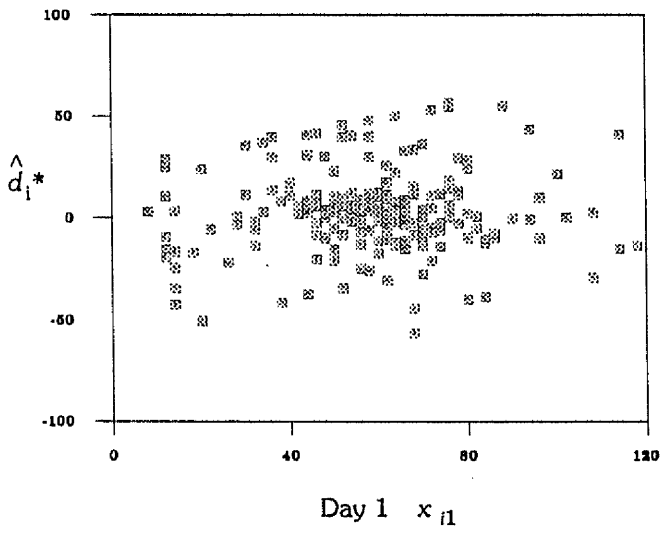


Figure 4.10 **Transformed $\hat{d}_i^*$ vs. Day 1 (Negative P wave Duration V1 in msecs).**

If we compare Figure 4.9 with Figure 4.10 it is clear that we have effectively removed the trend and indeed it appears that we may also be in the situation where $\tau = 0$.

The adapted model is now of the form

$$\hat{d}_i^* \sim N\left(0, 2\left(\sigma^2 + \tau^2 m_i\right)\right)$$

so that inferences about σ and τ can now be made in the manner of the previous sections.

Figure 4.11 provides approximate 99% and 95% confidence regions for the estimated values of $\hat{\sigma}$ and $\hat{\tau}$ for the negative P wave duration in lead V1 which are based on the adapted model. The maximum likelihood estimates are (14.0, 0.04). In fact, there is evidence to suggest that $\tau = 0$ is not all that implausible since this value lies inside the 95% confidence region in Figure 4.11.

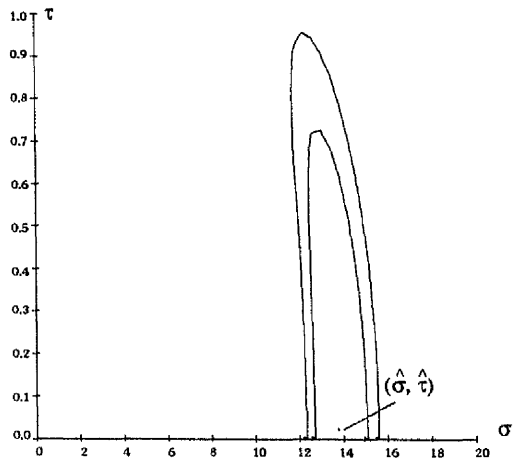


Figure 4.11 **Approximate 99% and 95% Confidence Regions for (σ,τ) (Negative P wave Duration in lead V1)**

4.4.6 Predicting the Day-to-day Variation.

Having obtained the maximum likelihood estimates for σ and τ, we may now proceed to predict the amount of variation which is likely to be present in a given day1 ECG measurement because this is typical of a one-off measurement.

Knowledge of the standard deviation of the differences between day 1 and day 2 measurements of various ECG parameters provides us with a method of assessing the amount of normal fluctuation

present in a single recorded value since all the natural variability is contained in σ, τ and μ_i for any patient i in the form

$$\sqrt{\text{Var(ECG Measurement)}} = \sqrt{\sigma^2 + \tau^2 \mu_i}$$

The MLE for the R wave duration in lead V1 is (0.01, 0.56). Therefore the estimated standard deviation of a second measurement of the R wave duration from an individual whose initial reading is around 30 msecs can be calculated as follows:

$$\text{Std. dev (RV1)} = \sqrt{(0.01^2 + 0.56^2 \times 30)}$$

$$= 3.07 \text{ msecs.}$$

An approximate prediction interval for the day 2 RV1 duration given that the day 1 reading was 30 msecs may be calculated as

$$30 \pm 1.96 \times 3.07 = (24, 36)$$

This would suggest that, for a day 1 RV1 duration of 30 msecs, we could expect a day 2 value of between 24 and 36 msecs. Any value outside this region is therefore likely to be viewed as significantly different from the day 1 reading. Figure 4.12 shows the range of values for the R wave duration in lead V1 which could be expected for typical one-off estimates of the measurement. In particular, it illustrates the prediction interval of (24, 36) msecs obtained for a recorded value of 30 msecs. The width of this prediction interval reflects the magnitude of the problem with which we are dealing.

Prediction bands for RV1 Duration

$$\mu \pm 1.96 \times \sqrt{(\hat{\sigma}^2 + \hat{\tau}^2 \mu)}$$

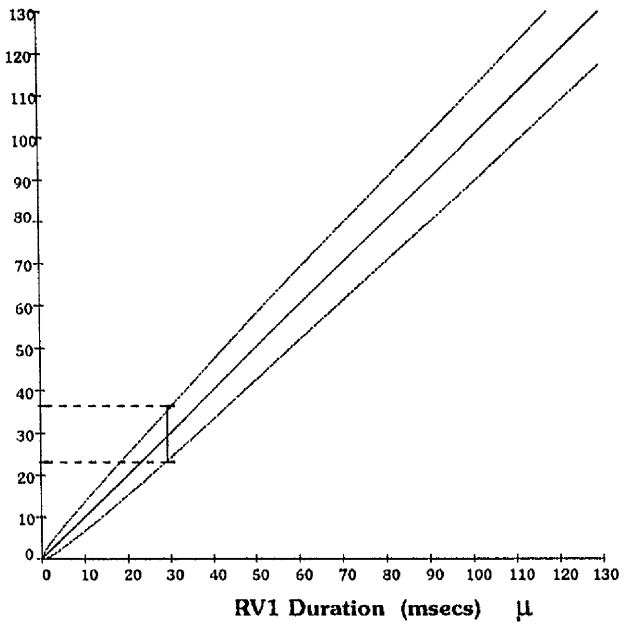


Figure 4.12 **Illustration of the range of plausible values for an observed RV1 Duration of 30 msecs.**

Similarly, prediction for the range of likely day 2 values for a given day 1 reading of the R wave amplitude in lead V5 is possible. For example, the standard deviation of the a second RV5 amplitude measurement given that the initial recorded value was 2500 μV can be constructed as follows:

i.e. $$\text{Std. dev (RV5)} = \sqrt{(28^2 + 5^2 \times 2500)}$$

$$= 252 \ \mu V.$$

The corresponding range of plausible day 2 values can be calculated as

$$2500 \pm 1.96 \times 252 = (2006, 2994)$$

The way in which the range of plausible day 2 values varies over the range of day 1 measurements is illustrated in Figure 4.13. Since day-to-day variation depends on the magnitude of the ECG variable under consideration (in this particular case, RV5), the standard deviation of the differences for high day 1 measurements will be greater than the standard deviation for lower values. This will in turn result in wider prediction bands for high day 1 readings.

Prediction bands for RV5 Amplitude

$$\mu \pm 1.96 \times \sqrt{(\hat{\sigma}^2 + \hat{\tau}^2 \mu)}$$

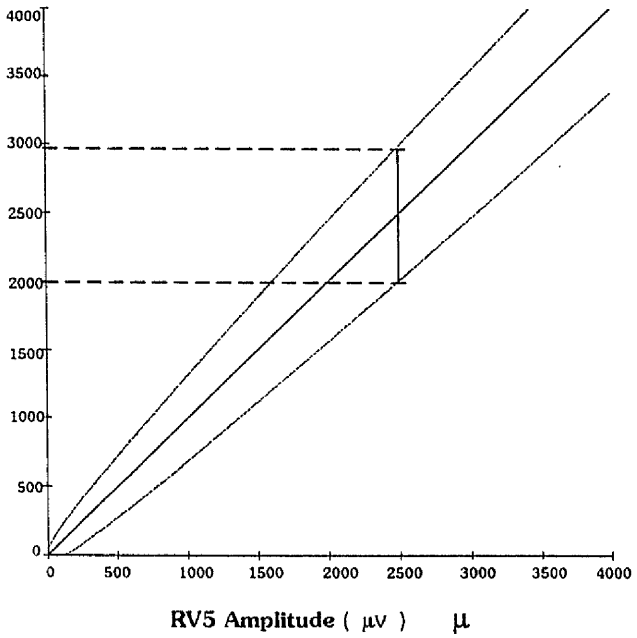


Figure 4.13 **Illustration of the range of plausible values for an observed RV5 Amplitude of 2500μV.**

4.5 THE EFFECTS OF 'STRANGE' OBSERVATIONS.

4.5.1 The Presence and Detection of Outliers.

Any moderately sized sample taken from a general population is likely to contain a few values which are surprisingly far away from the main group. Reasons for this might be natural variability within the sample or some form of technical malfunction leading to an aberrant value. The problem here lies in deciding whether to disregard these values completely or to include them together with the remaining observations when interpreting the data since any analysis might be sensitive to the presence of these possible outliers.

A simple plot of the square root of the day 1 measurement of the R wave amplitude in lead II versus the difference of the square root of

the day 1 and the day 2 measurements has been provided (Fig 4.14). Clearly there is one extreme observation where the difference appears to be much larger than would have been anticipated given the initial measurement. However, the difference may be a true one so that information might be lost if this reading were excluded from the analysis.

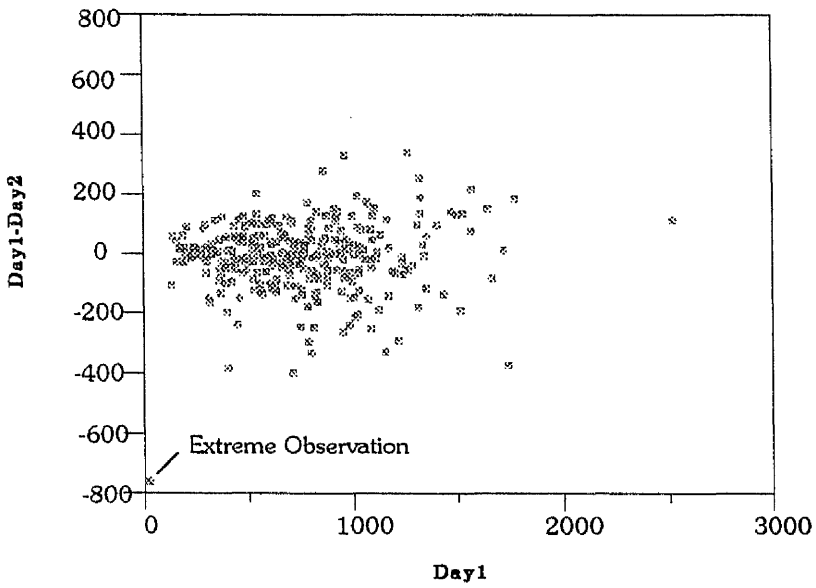


Figure 4.14 RII Amplitude (Day1 vs. Day1-Day2)

This extreme observation may have arisen from random measurement error or from other sources that are of no particular relevance to the study (Barnett and Lewis, 1977). The effect however on the parameter estimates of σ and τ may be out of all proportion so one solution is to dampen the effect of the possible outlier on the estimates using the idea of an influence function. The main aim of this is to reduce the effect of the possible outlier on the likelihood function (Hampel, 1974).

4.5.2 The Use of Influence Functions to dampen the effects of potential outliers.

Complete rejection of possible outliers is unsatisfactory since this may contribute to underestimation of the parameters of interest. Alternatively, treating extreme observations in the same way as the rest of the sample may provide overestimates of the parameters out of proportion to that merited by one or two 'wild observations'.

One solution is to introduce an influence function which dampens the effect of observations which are 'far from the rest' so that they do not distort parameter estimates too much. In this way, all available and relevant data will be included when making inferences about σ and τ although observations which appear to be exerting undue influence on the underlying model will be considered with caution.

The influence function measures the effect that a single observation may have on the criterion (i.e. likelihood) function which has been used to estimate the value(s) of the parameter(s) which maximise the likelihood or profile likelihood function. Naturally, outliers which offer 'substantial contributions' to parameter estimates will significantly impair the performance of the model unless a suitable influence function is used.

Consider a sample of real-valued observations $x_1,...,x_n$. Two obvious estimators of the population 'centre' θ are the sample mean $\hat{\theta}_A(\underline{x})$ and the sample median $\hat{\theta}_B(\underline{x})$.

$$\hat{\theta}_A(\underline{x}) = \frac{\sum_{i=1}^{n} x_i}{n} \quad ;$$

$$\hat{\theta}_B(\underline{x}) = \text{sample median of } (x_1,\ldots,x_n) = \begin{cases} \left(x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)} \right)/2 & \text{if } n \text{ is even} \\ x_{\left(\frac{n+1}{2}\right)} & \text{if } n \text{ is odd} \end{cases}$$

Note that $x_{(\,)}$ denote order statistics.

An additional observation x will alter these estimates in the following ways:

$$\hat{\theta}_{A'}(\underline{x}) = \frac{n}{n+1}\hat{\theta}_A(\underline{x}) + \frac{1}{n+1}x;$$

$$\hat{\theta}_{B'}(\underline{x}) = \text{sample median of } (x_1,\ldots,x_n,x) = \begin{cases} x_{\left(\frac{n}{2}+1\right)} & \text{if } n \text{ were even} \\ \left(x_{\left(\frac{n+1}{2}\right)} + x_{\left(\frac{n+1}{2}+1\right)} \right)/2 & \text{if } n \text{ were odd} \end{cases}$$

in the (new) order statistic of the (n+1) observations.

Consider now the effect of the magnitude of x on the value of the parameter estimate. Here we plot x vs. the difference in the estimate of the sample mean (or median) observed when the additional observation x is introduced, i.e.

$$\hat{\theta}_{A'}(\underline{x}) - \hat{\theta}_A(\underline{x}) \text{ vs. } x \quad \text{or} \quad \hat{\theta}_{B'}(\underline{x}) - \hat{\theta}_B(\underline{x}) \text{ vs. } x$$

In the situation where the location of a particular population is estimated by its sample mean $\hat{\theta}_A(\underline{x})$, the influence that an additional observation x will have on the estimate can be seen in Figure 4.15.

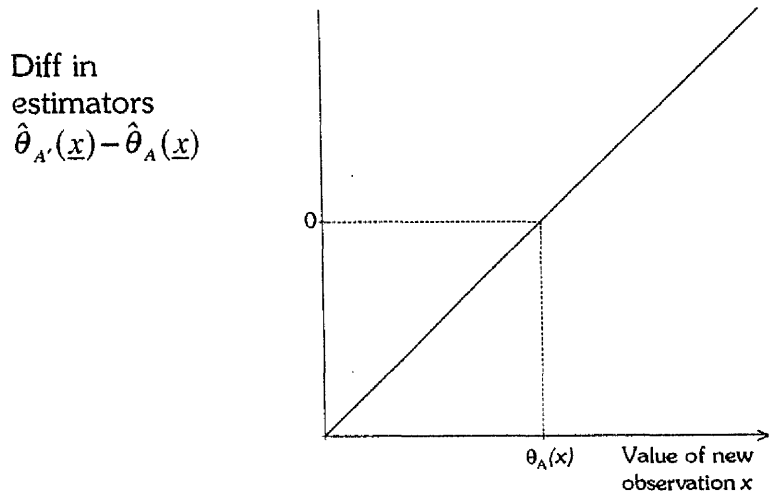


Figure 4.15 **Influence that x has on the sample mean.**

If $x=\hat{\theta}_A(\underline{x})$, then the influence on the estimate is zero, but as x increases, the difference between the estimates increases producing an unbounded influence function.

However, using the sample median as an estimate will produce a bounded influence function which can be seen in Figure 4.16.

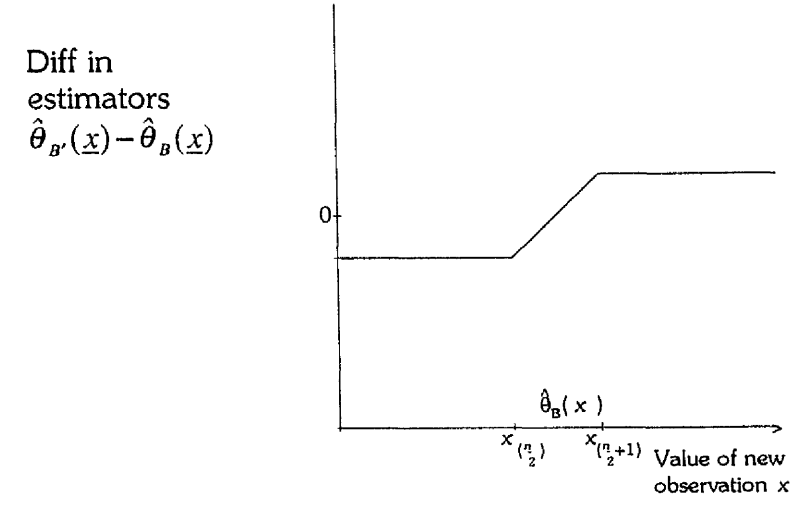


Figure 4.16 **Influence that x has on the sample median in the case where n is even.**

Estimating the location of the population mean by the sample median is therefore more robust than using the sample mean, offering more protection against outliers. This is largely due to the fact that the inclusion of an extra observation will influence the sample median only slightly whereas the sample mean may be altered quite dramatically depending on the magnitude of the new observation relative to the rest of the sample.

Due to our approximate Profile Likelihood approach, interest lies primarily in the recorded differences between day 1 and day 2 observations. Therefore a method of coping with 'large' or 'influential' values is required since extreme differences will tend to result in the over-estimation of σ and τ in the model for day-to-day variation.

If we consider the form of the profile likelihood (i.e. $\hat{pl}(\sigma,\tau;\underline{x}) = l(\sigma,\tau,m_i;\underline{x})$) each of the data points contributes a function of $\frac{d_i^2}{\sigma^2+\tau^2 m_i}$ which is clearly unbounded as a function of d_i.

Plotting the $\frac{d_i}{\sqrt{\sigma^2+\tau^2 m_i}}$ against their squared value for RV5 illustrates several observations which are far away from the majority of the data (see Figure 4.17).

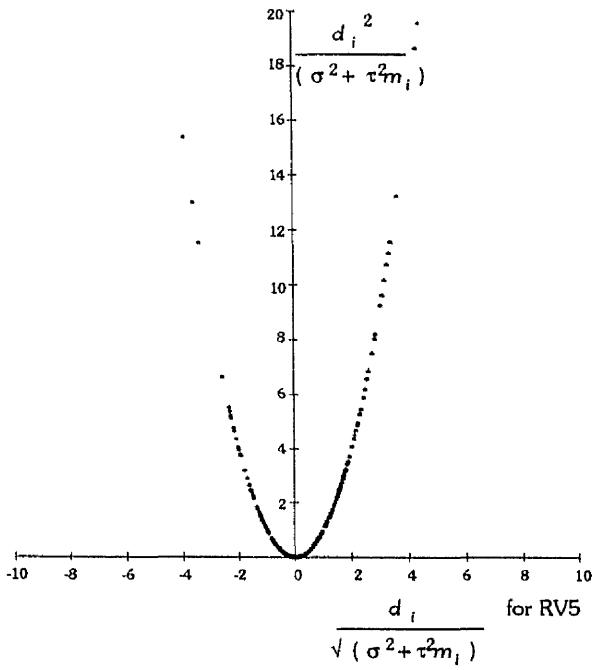


Figure 4.17 **Plot of the unbounded influence function for RV5 Amplitude**

The observations which are far away from the majority of the data will have a greater influence on the maximum likelihood estimates $(\hat{\sigma},\hat{\tau})$ than is reasonable and will tend to overestimate σ and τ.

One way to remove the 'unboundedness' is simply to cut off the function at a (rather arbitrary) point and restrict the influence of any larger observation to this cut-off value.

Consider

$$\vartheta(d_i)=\begin{cases}\dfrac{d_i^2}{(\sigma^2+\tau^2 m_i)} & \text{if } \dfrac{d_i}{\sqrt{(\sigma^2+\tau^2 m_i)}}\le k \\ k^2 & \text{if } \dfrac{d_i}{\sqrt{(\sigma^2+\tau^2 m_i)}}> k\end{cases}$$

This is an example of a bounded influence function based on the differences d_i although a somewhat arbitrary choice for k is required.

Since $\frac{d_i^2}{\sigma^2+\tau^2 m_i}$ appears in the Profile Likelihood (i.e. the criterion function), $\frac{d_i^2}{\sigma^2+\tau^2 m_i}$ will be replaced by $\vartheta(d_i)$. Values of $\left|\frac{d_i}{\sqrt{\sigma^2+\tau^2 m_i}}\right|$ which are of a greater magnitude than $|k|$ will be 'brought in' towards the bulk of the sample otherwise they will have the usual 'least squares influence' on the likelihood.

The obvious problem of the somewhat arbitrary choice of k might be solved by a robust estimate for k such as $k=\frac{3}{2}\,\mathrm{IQR}\left(\frac{d_i}{\sqrt{(\hat{\sigma}^2+\hat{\tau}^2 m_i)}}\right)$. The reasons for this choice are that the IQR (Inter-Quartile Range) will not itself be influenced by one or two outliers and the multiple $\frac{3}{2}$ times the IQR might roughly correspond to two standard deviations of the population. Hence only observations which are further than approximately two standard deviations from zero (the assumed mean of the d_i) will be reduced in influence.

The approximate profile likelihood will therefore become

$$\hat{pl}(\sigma,\tau;\underline{x})^* = -\sum_{i=1}^{n}\log(\sigma^2+\tau^2 m_i) - \sum_{i=1}^{n}\frac{\vartheta(d_i)}{4}$$

where $m_i=\frac{x_{i1}+x_{i2}}{2}$ and $d_i=(x_{i1}-x_{i2})$.

Plotting the $\frac{d_i}{\sqrt{\sigma^2+\tau^2 m_i}}$ against their squared value for RV5 illustrates several observations far away from the majority of the data (see Figure 4.17). Figure 4.18 demonstrates how consecutive

recordings which exhibit differences in ECG measurements lying outside the range $(-k,k)$ will be 'brought in' towards the remainder of the sample so that the estimated values of σ and τ will not be influenced unduly.

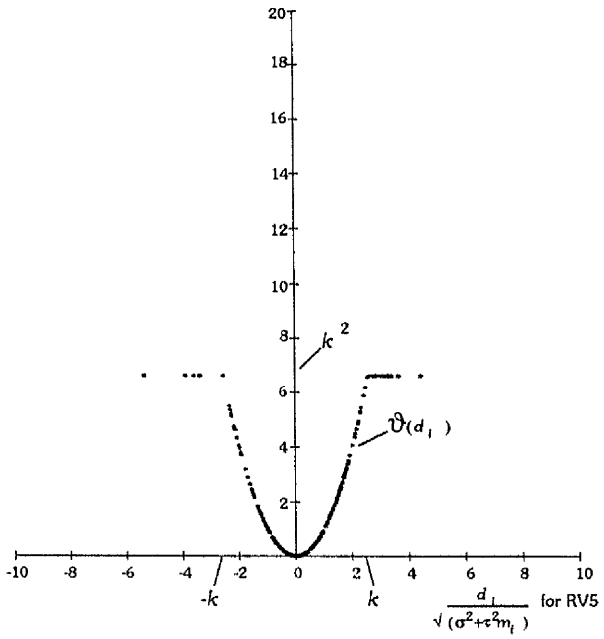


Figure 4.18 **Example of a bounded influence function for RV5 Amplitude**

The surface plot and the contour plot of the approximate log-likelihood for the RV5 amplitude when influential observations are treated in the way described can be seen in Figures 4.19 and 4.20.

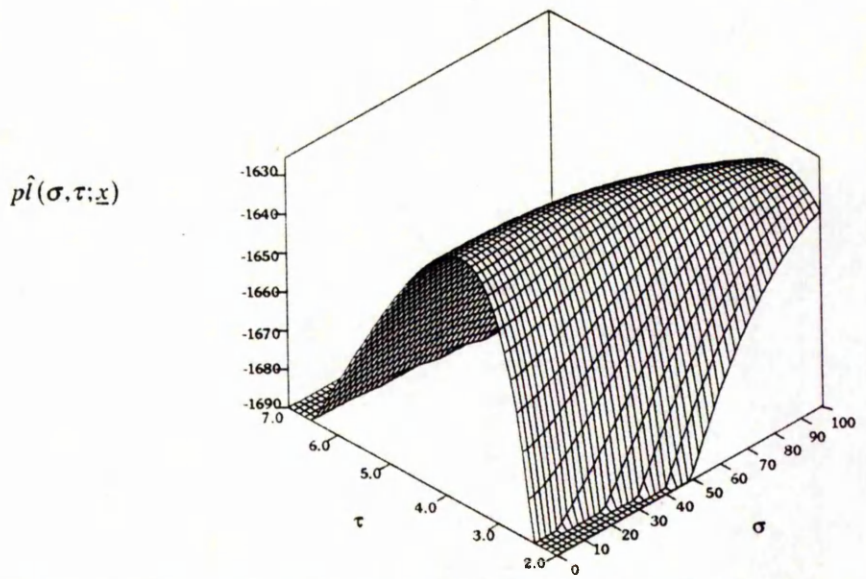


Figure 4.19 **Surface plot of the Approximate log-likelihood function for RV5 Amplitude (taking account of influential observations) (cf. Figure 4.7)**

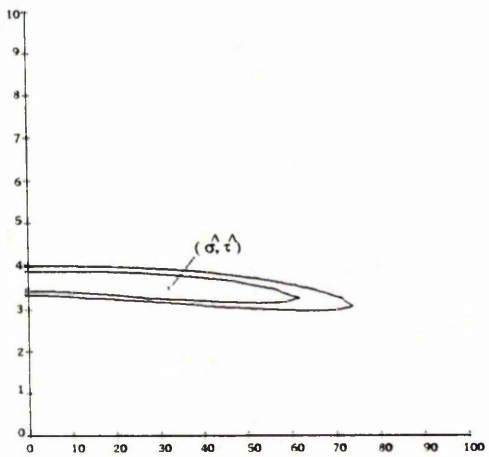


Figure 4.20 **Approximate 99% and 95% Confidence Regions for (σ,τ) (RV5 Amplitude) (taking account of influential observations) (cf. Figure 4.8)**

Using this modification, the MLE for (σ,τ) is now (32, 3.7) for RV5 compared to the previous estimate of (28, 5.0). Thus, for an RV5 reading of 2500 μV, the standard deviation of the day-to-day variation will be estimated at around 189 μV for RV5 using the robust procedure compared with a value of 252 μV estimated from

the full data set including any potential outliers. The resulting prediction range for the RV5 amplitude will be (2130, 2870) compared to (2006, 2994). Treating the outliers as normal observations caused an increased estimate of τ and hence an inflated estimate of the standard deviation. In turn, this produces a wider range of plausible values for RV5. Using the robust procedure described has therefore solved the problem of deciding what rules to apply when attempting to recognise and perhaps discard potential outliers. Instead, all observations are included when inferences are being made about (σ, τ) although 'strange' observations will have only a limited influence on the estimates.

4.6 ESTIMATES OF DAY-TO-DAY VARIABILITY FOR THE COMPLETE SET OF ECG VARIABLES.

The methods which have been described in this chapter were applied to the complete set of ECG variables. Due to limitations of space, the findings for two subsets of amplitudes and durations (R wave and ST segment) will be presented.

Section 4.6.1 provides the results for the R wave amplitudes and Section 4.6.2 examines the R wave durations. In both cases the model described in section 4.4.1 has been used. Results relating to the ST amplitudes and the ST durations (Sections 4.6.3 and 4.6.4 respectively) are based on the alternative model which has been described in Section 4.4.5.

4.6.1 Results for the R wave amplitudes.

The following inferences about the R wave amplitudes have been based on the model described in section 4.4.1. This model was based on the assumption that there was no unusual trend in the day 1 and day 2 recordings and that the observed difference between the measurements may or may not depend on the day 1 reading,

i.e.

$$\text{DAY 1} \qquad x_1 = \mu_1 + \varepsilon_1$$
$$\text{DAY 2} \qquad x_2 = \mu_2 + \varepsilon_2$$

where $\varepsilon_1 \sim N(0, \sigma^2 + \tau^2\mu)$, $\varepsilon_2 \sim N(0, \sigma^2 + \tau^2\mu)$ and $d \sim N(0, 2(\sigma^2 + \tau^2\mu))$, where μ is the mean of the day 1 and the day 2 readings and $\tau = 0$ or $\tau > 0$ as appropriate. All the ECG variables in this subset are adequately described by this model.

The maximum likelihood estimates for σ and τ are provided in Table 4.2. In addition are the results of two likelihood ratio tests, the first examining the null hypothesis that $\sigma = 0$ and the second looking at the possibility of $\tau = 0$. Non-significant results respectively suggest that $\sigma = 0$ or $\tau = 0$ are reasonable. The test of $\tau = 0$ vs $\tau > 0$ is of greater interest than the test of $\sigma = 0$ vs $\sigma > 0$ since if there is insufficient evidence to reject the null hypothesis that $\tau = 0$ then the computation becomes less awkward due to the elimination of nuisance parameters μ_i. On the other hand, if we can assume that $\sigma = 0$ then we are eliminating only the baseline variability which does not necessarily simplify subsequent calculations to any great extent. The median value for each day 1 amplitude has been given (x_m) along with the ratio of the estimated standard deviation calculated at the sample median, to the sample median (σ_m^* / x_m) multiplied by

100. This provides a method of comparing the relative amounts of day-to-day variability in the different leads. The final column provides the 95% prediction range of measurements likely to be seen at a typical value of the amplitude (in this case, the sample median x_m)

Lead	$(\hat{\sigma},\hat{\tau})$	Is $\sigma>0$?	Is $\tau>0$?	Sample Median x_m	% Relative s.d. at x_m $\frac{\sigma_m^*}{x_m}\times 100$	Prediction Range at x_m
I	(35.2,1.68)	yes	yes	812	7.3	(696,928)
II	(24.0,1.92)	yes	yes	721	7.9	(610,832)
III	(1.6,2.50)	no	yes	238	16.2	(162,314)
aVR	(0.8,1.20)	no	yes	74	14.0	(54,94)
aVL	(12.8,2.16)	yes	yes	585	9.2	(480,690)
aVF	(1.2,2.48)	no	yes	499	11.1	(390,607)
V1	(0.4,1.52)	no	yes	191	11.0	(150,232)
V2	(0.4,2.64)	no	yes	448	12.5	(338,558)
V3	(0.8,3.92)	no	yes	810	13.8	(591,1029)
V4	(0.8,4.20)	no	yes	1368	11.4	(1063,1672)
V5	(32.0,3.72)	yes	yes	1466	10.0	(1180,1752)
V6	(0.8,3.72)	no	yes	1270	10.4	(1010,1530)

Table 4.2 Table of results for the amplitude of the R wave (μV) based on 295 patients

Day-to-day variation in all twelve leads does appear to depend on the magnitude of the ECG measurement since in each case τ is significantly greater than zero.

A plot of the difference in the day 1 and day 2 readings of RV2 can be seen in Figure 4.21. From Table 4.2 it has been shown that the day-to-day variation in the R wave amplitude in lead V2 depends on the day 1 reading. The data points in Figure 4.21 give a slight impression of a 'wedge-shape' and demonstrate that the difference between day 1 and day 2 measurements increases as the initial

reading increases, i.e. $\tau > 0$ in our model for day-to-day variation. The percentage relative standard deviation at a typical day 1 measurement of RV2 is 12.5%. The lines which have been superimposed on top of the data points in Figure 4.21 represent the expected mean value of the d_i plus or minus two times the estimated standard deviation $\left(\text{i.e. } 0 \pm 2\left(\sqrt{2\left(\hat{\sigma}^2 + \hat{\tau}^2 x_{i1}\right)}\right)\right)$. If the differences are Normally distributed (and we expect that they are since much of the variation between patients has been removed, leaving only day-to-day variability) then we expect that 95% of the differences will lie between these limits.

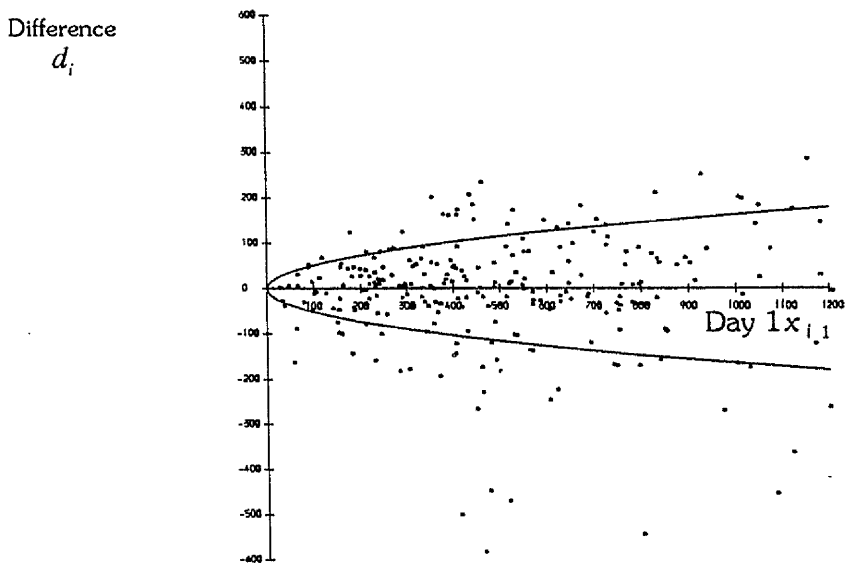


Figure 4.21 **Plot of Day 1 vs. Day1-Day2 (RV2 Amplitude)**

There appears to be a larger amount of baseline variability in day-to-day readings of the R wave amplitude in lead I. The MLE for RI is (35.2, 1.68) compared to (0.4, 2.64) for RV2. This results in larger amounts of day-to-day variability for low day 1 readings and proportionately smaller amounts of day-to-day variability for higher day 1 measurements (see Figure 4.22). Correspondingly, the

percentage relative standard deviation for a typical day 1 value is 7.3%.

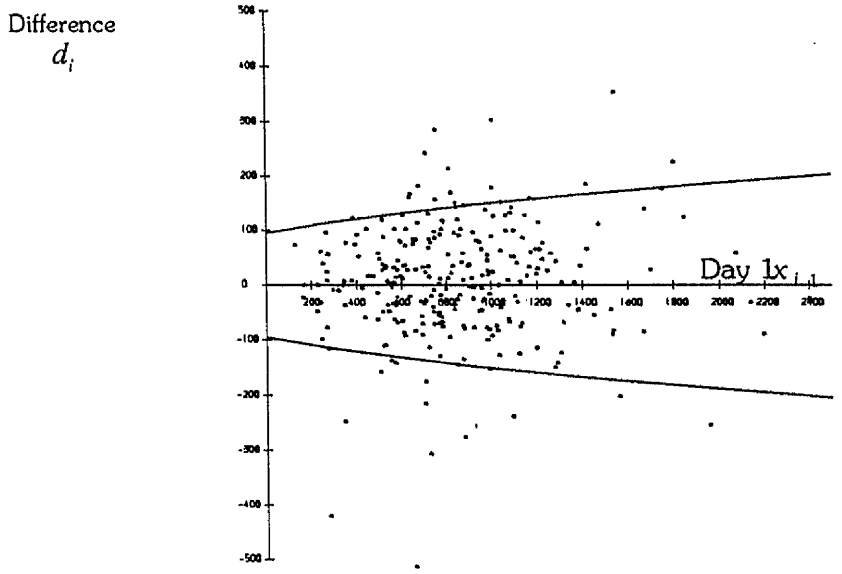


Figure 4.22 **Plot of Day 1 vs. Day1-Day2 (RI Amplitude)**

4.6.2 Results for the R wave durations.

Results for the R wave durations in each of the twelve leads are provided in Table 4.3. Again, the estimates of σ and τ are based on the model described in section 4.4.1. It is interesting to note that in each lead except one (lead V1), $\tau > 0$ indicating that the day-to-day variation in the duration of the R wave in the majority of the twelve leads depends, at least to some extent, on the magnitude of the particular duration of interest. Furthermore, in leads II, aVR, aVL, V2, V4 and V6 the day-to-day variation may be expressed solely as a multiple of the magnitude since σ can be ignored. The ratio of the estimated standard deviation at the sample median to the sample median itself is provided as a means of determining the relative amount of variability in each lead.

Lead	$(\hat{\sigma},\hat{\tau})$	Is $\sigma>0$?	Is $\tau>0$?	Sample median x_m	%Relative s.d. at x_m $\frac{\sigma^*_m}{x_m}\times 100$	Prediction Range at x_m
I	(1.4,0.3)	yes	yes	59	4.2	(54,64)
II	(0.1,0.3)	no	yes	57	4.2	(52,62)
III	(1.9,0.4)	yes	yes	36	8.0	(30,42)
aVR	(0.1,0.8)	no	yes	21	17.0	(14,28)
aVL	(0.1,0.4)	no	yes	54	5.2	(49,59)
aVF	(1.0,0.4)	yes	yes	47	5.9	(42,52)
V1	(1.7,0.1)	yes	no	27	6.3	(24,30)
V2	(0.6,0.3)	no	yes	33	5.6	(29,37)
V3	(1.0,0.3)	yes	yes	42	5.5	(37,47)
V4	(0.1,0.3)	no	yes	46	4.2	(42,50)
V5	(1.0,0.2)	yes	yes	48	4.0	(44,52)
V6	(0.2,0.3)	no	yes	58	3.7	(54,62)

Table 4.3 **Table of results for the duration of the R wave (msecs) based on 295 patients.**

Although day-to-day variation depends on the magnitude of the ECG measurement of interest for most of the R wave durations, there are some differences observed in the ratios of σ^*_m to x_m. For example, the percentage relative standard deviation at the sample median of the R wave duration in lead I is around 4.2% (Figure 4.23) whereas it is four times this amount (17%) in lead aVR (Figure 4.24).

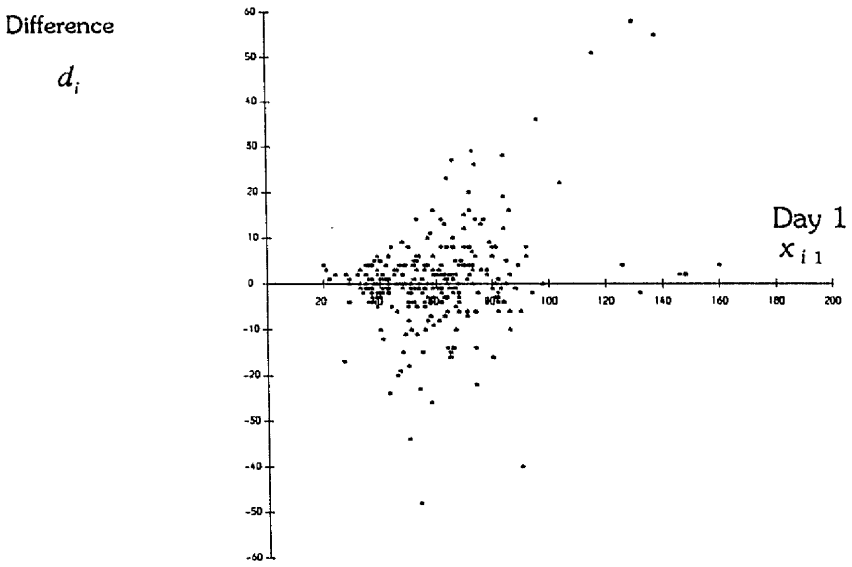


Figure 4.23 Plot of Day 1 vs. Day1-Day2 (RI Duration)

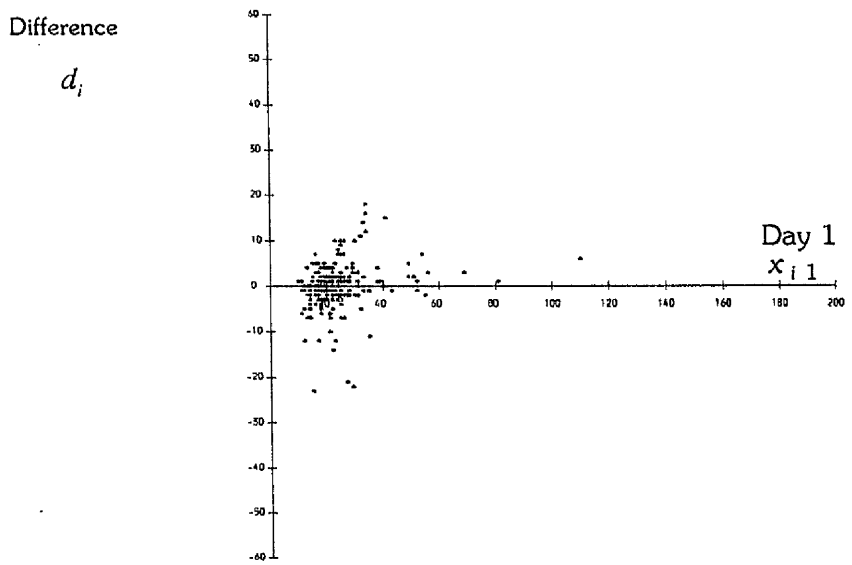


Figure 4.24 Plot of Day 1 vs. Day1-Day2 (RaVR Duration)

Generally, the day-to-day variation appears to be lower for the R wave durations than for the R wave amplitudes. The percentage relative standard deviations range from 7.3% (RI) to 16.2% (RIII) for the R wave amplitudes. Ignoring the duration of the R wave in lead aVR, the corresponding range for the R wave durations is 3.7% (lead

V6) to 8% (lead III). Day-to-day measurements of the R wave amplitudes are therefore liable to more variation than day-to-day readings of the R wave durations.

4.6.3 Results for the ST segment amplitudes.

In general for the ST segment data the 'regression to the mean' phenomenon was found to appear in all leads. Thus the alternative model which takes into account any evidence of a trend between the day 1 and day 2 observations was used to produce maximum likelihood estimates of σ and τ for the ST segment amplitudes (see section 4.4.5), i.e.

$$\text{DAY1} \qquad X_{i1} = \mu_i + \varepsilon_{i1}$$

$$\text{DAY2} \qquad X_{i1} = \mu_i - \lambda(\mu_i - \bar{\mu}) + \varepsilon_{i2}$$

where $\varepsilon_{ij} \sim \mathrm{N}\left(0, \left(\sigma^2 + \tau^2 \mu_i\right)\right)$ for $i = 1,\ldots,n;\ j = 1,2$ and $\lambda > 0$ is a measure of the rate of 'regression to the mean'.

Lead	$\hat{\lambda}$	Is $\lambda > 0$?	$(\hat{\sigma}, \hat{\tau})$	Is $\sigma > 0$?	Is $\tau > 0$?	Sample median x_m	%Relative s.d. at x_m $\frac{\sigma^*_m}{x_m} \times 100$	Prediction Range at x_m
I	0.20	yes	(8.0,0.28)	yes	yes	18	45	(2,33)
II	0.16	yes	(9.6,0.18)	yes	yes	23	42	(4,41)
III	0.20	yes	(8.0,0.16)	yes	yes	20	40	(4,35)
aVR	0.17	yes	(6.4,0.5)	yes	yes	20	34	(7,33)
aVL	0.23	yes	(4.8,0.52)	yes	yes	14	37	(4,24)
aVF	0.21	yes	(4.2,0.45)	yes	yes	16	29	(7,25)
V1	0.15	yes	(6.4,1.0)	yes	yes	52	18	(33,71)
V2	0.12	yes	(5.6,0.96)	yes	yes	90	12	(69,111)
V3	0.22	yes	(11.2,0.64)	yes	yes	66	19	(42,90)
V4	0.30	yes	(8.0,0.52)	yes	yes	37	23	(21,54)
V5	0.29	yes	(7.2,0.4)	yes	yes	29	26	(14,43)
V6	0.31	yes	(10.4,0.06)	yes	no	26	43	(4,44)

Table 4.4 **Table of results for the amplitude of the ST segment (μV) based on 295 patients.**

In Table 4.4 the value $\hat{\lambda}$ is an estimate of the rate of 'regression to the mean' and in each case it is significantly different from zero.

In eleven of the twelve leads, day-to-day variation depends on the magnitude of the ST amplitude. In the remaining lead (V6) the variability may be expressed as a constant since τ is effectively zero. Figure 4.25 illustrates that the amount of day-to-day variation in the standardised differences of lead V6 is reasonably constant whereas Figure 4.26 demonstrates that the variation in the $\hat{d}_i^*$ of lead V1 increases as the day 1 magnitude x_m increases.

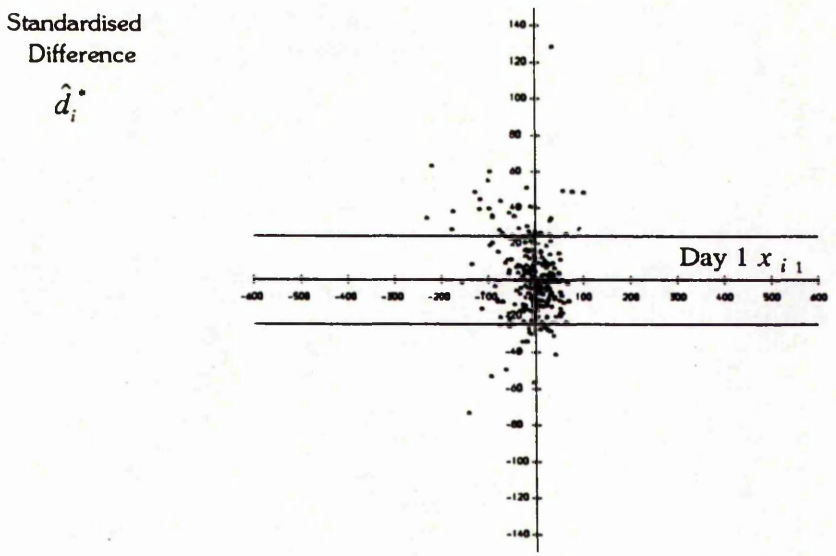


Figure 4.25 **Plot of Day 1 vs. $\hat{d}_i^*$ (ST Amplitude Lead V6)**

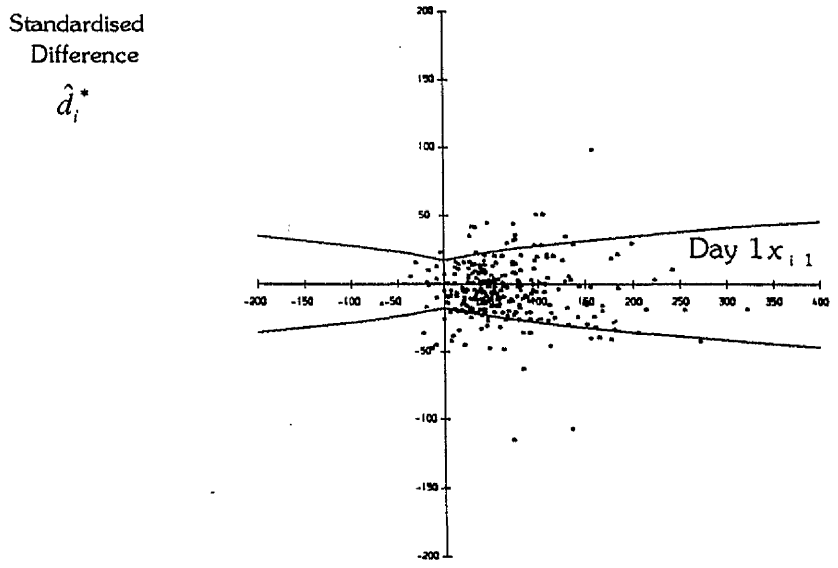


Figure 4.26 **Plot of Day 1 vs. $\hat{d}_i^*$ (ST Amplitude Lead V1)**

Approximately 95% of the differences $\hat{d}_i^*$ lie within the boundary lines which have been superimposed on the plots. These boundaries are straight lines in Figure 4.25 since $\tau = 0$ and increase as $|x|$ increases in Figure 4.26 due to the fact that $\tau > 0$.

The percentage relative standard deviations at the sample median values for the ST amplitudes ranged from 12% (lead V2) to 45% (lead I). In four of the leads (I, II, III and V6) these relative standard deviations are in excess of 40% which gives considerable cause for concern with the reliability of these measurements.

4.6.4 Results for the ST segment durations.

The alternative model which takes account of a trend from the day 1 to the day 2 recordings was also used to provide maximum likelihood estimates of σ and τ for the ST segment durations (see Table 4.5). It should be noted that the ST segment duration is a

somewhat arbitrary interval which is not used clinically, but serves to illustrate the methods which have been used.

Lead	$\hat{\lambda}$	Is $\lambda > 0$?	$(\hat{\sigma}, \hat{\tau})$	Is $\sigma > 0$?	Is $\tau > 0$?	Sample median x_m	% Relative s.d at x_m $\frac{\sigma^*_m}{x_m} \times 100$	Prediction Range at x_m
I	0.56	yes	(19.2,0.0)	yes	no	120	16	(82,158)
II	0.59	yes	(18.0,0.0)	yes	no	120	15	(85,155)
III	0.57	yes	(19.2,0.0)	yes	no	122	16	(84,160)
aVR	0.56	yes	(18.4,0.2)	yes	no	120	15	(84,156)
aVL	0.62	yes	(19.6,0.0)	yes	no	120	16	(82,158)
aVF	0.58	yes	(18.0,0.0)	yes	no	126	14	(91,161)
V1	0.46	yes	(16.8,0.2)	yes	no	115	15	(82,148)
V2	0.37	yes	(12.4,0.0)	yes	no	110	11	(86,134)
V3	0.43	yes	(12.6,0.0)	yes	no	108	12	(83,133)
V4	0.45	yes	(12.6,0.0)	yes	no	114	11	(89,139)
V5	0.43	yes	(14.0,0.0)	yes	no	116	12	(89,143)
V6	0.45	yes	(14.8,0.2)	yes	no	120	13	(91,149)

Table 4.5 Table of results for the duration of the ST segment (msecs) based on 295 patients.

For the ST durations, the estimated standard deviation at a typical value, x_m, of each of the leads I to aVF appears reasonably constant (i.e. between 18 and 19.6). A plot of the day 1 ST duration in lead I vs. the standardised difference $\hat{d}_i^*$ can be seen in Figure 4.27.

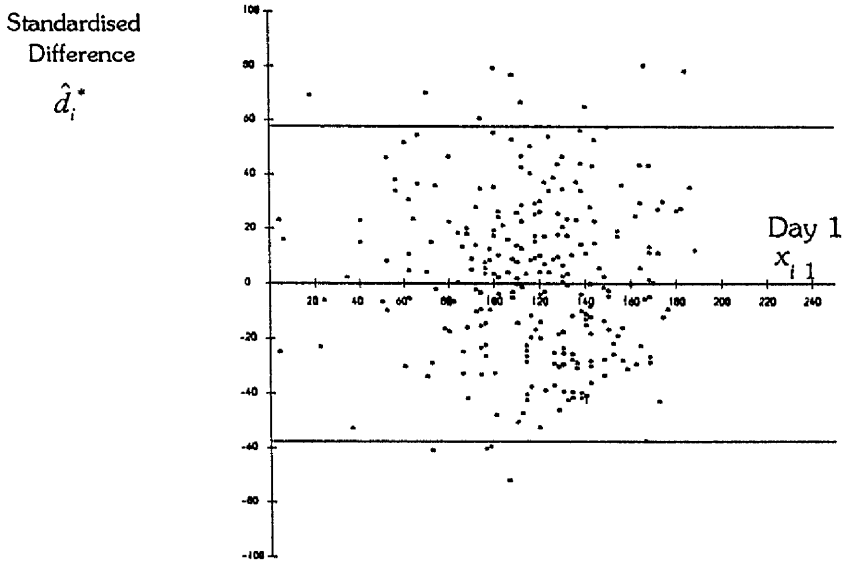


Figure 4.27 **Plot of Day 1 vs. $\hat{d}_i^*$ (ST Duration Lead I)**

There appears to be slightly less variability in the $\hat{d}_i^*$ of the durations of the ST segment in the chest leads V1 to V6. The estimated standard deviation of $\hat{d}_i^*$ at the sample median ranges from 12.3 for lead V2 to 16.9 for lead V1 (see Table 4.5). Figure 4.28 illustrates that most of the $\hat{d}_i^*$ for lead V4 lie in a narrower region than that for lead I.

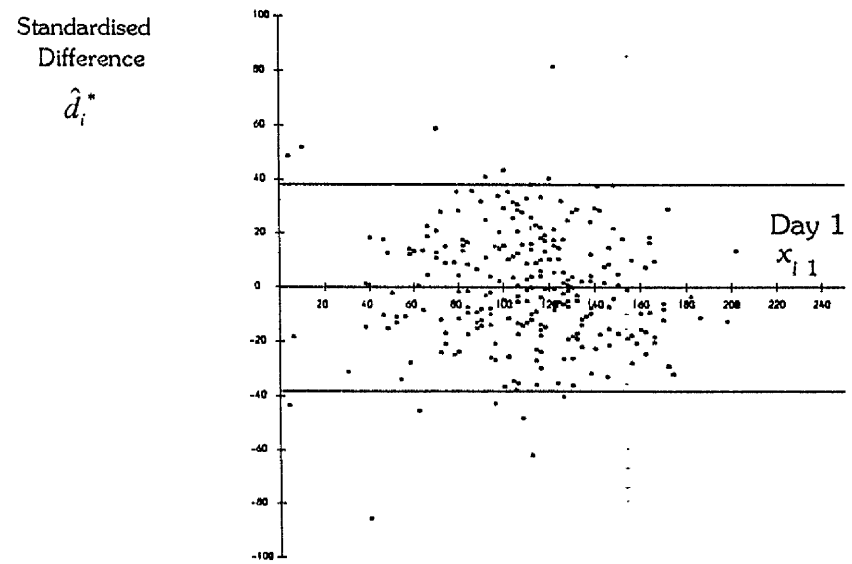


Figure 4.28 **Plot of Day 1 vs. $\hat{d}_i^*$ (ST Duration Lead V4)**

4.7 SUMMARY.

The behaviour of ECG measurements which have been taken from replicate recordings has been investigated. It has been demonstrated that the observed differences between consecutive recordings (often only 24 hours apart) in many ECG variables are by no means negligible and such discrepancies contribute to lack of repeatability in the diagnostic process.

The standard model suggested for the day-to-day variation in any ECG measurement was based on the assumption that the differences between day 1 and day 2 readings of selected ECG variables demonstrated no apparent trend apart from a possible dependence on the day 1 measurement. However, such a model appeared to be inadequate in certain cases (notably when the P wave and the ST segment were considered) and hence two models for day-to-day variation were investigated.

The initial model was based on the assumption that the expected difference between day 1 and day 2 observations of the ECG measurement of interest may or may not depend, in some simple way, on the ECG variable. The second model allowed for a trend between the day 1 and day 2 measurements which was suggested when the data were explored. Both models considered whether the amount of day-to-day variation was dependent on the magnitude of the particular ECG measurement of interest by introducing two unknown parameters, σ which measures any baseline variability and τ which assesses the relative increase of the variance with the magnitude of the ECG variable.

The application of these techniques to particular cases permitted the estimation of the amount of day-to-day variation in many ECG variables and some of these estimates have been presented. Likelihood ratio tests were performed in each case to determine if the variability depended on the magnitude of the ECG measurement, i.e. whether $\tau = 0$ was reasonable.

Tables of results for the R wave amplitudes and durations have been provided and the estimates $(\hat{\sigma}, \hat{\tau})$ are based on the initial model which is described in 4.4.1. The percentage relative standard deviation at the sample median ranged from 7.3% (in lead I) to 17.2% (in lead III) in the case of the R wave amplitudes and from 3.7% (lead V6) to 17% (lead aVR) for the R wave durations. With the exception of lead aVR, the duration of the R wave seems to vary less from day to day than do the amplitudes.

Inferences about the ST segment amplitudes and durations have been based on the model described in section 4.4.5 and these results are also provided. The percentage relative standard deviations for the ST segment amplitudes are substantial, ranging from 12% (lead V2) to 45% (lead I). In contrast, these percentages are smaller for the ST segment durations, the maximum being 16% for leads I, III and aVL.

In general, the percentage differences in the day to day readings of the ST segment (i.e. the ST amplitude and the ST duration) are of a greater magnitude than the day-to-day differences observed in the R wave measurements although in both cases there is less variability in the day to day measurement of the durations.

In conclusion, the relationship between day-to-day variation and the magnitude of the ECG variable of interest must be considered when taking into account the repeatability of the ECG process.

CHAPTER FIVE:

SOURCES OF REPEAT VARIABILITY IN THE ECG:

Diagnostic Thresholds.

5.1 THE STATUS QUO.

At present, the Glasgow Program relies on a deterministic approach in order to diagnose ECG abnormalities. The reasons for adopting such an approach have been outlined in Chapter One. However, certain problems have been encountered when examining the repeatability of the procedure. Such problems have undoubtedly arisen

1) as a result of the presence of discrete thresholds in the diagnostic program which are often crossed between successive recordings on the same patient and
2) from the natural variability in ECG measurements which exists from day to day or indeed from recording to recording.

Methods of minimising the effects of this lack of repeatability on the diagnostic decisions are required and the following chapters examine various techniques which will attempt to achieve this.

5.2 INTRODUCTION.

Replacing the discrete age and sex categorised upper limits of 'healthiness' for various ECG variables with continuous analogues to some extent ameliorates the problem of small measurement changes contributing to conflicting diagnoses when, for instance, age category alters between successive recordings. However, there remains the much more common problem of the ability to cope with small measurement changes which cause a discrete threshold to be crossed between two consecutive recordings even though there is no considerable change in overall ECG appearance. For example, if a Caucasian male aged 30 has an RV5 amplitude of 3.1mV on a first recording and 3.2mV on a subsequent recording these values lie on either side of the continuous upper limit of 'healthiness' previously derived for RV5. It is reasonable to assume that there is no clinically significant difference between these readings although, in practice, the first recording would have scored 0 points on the 'diagnostic index', while the second would score 2 points, the latter contributing to a possible final diagnosis of Left Ventricular Hypertrophy (see later).

It is therefore important to be able to control the extent to which a measurement is to be considered abnormal. One approach is to develop a smoothing function where the number of points allocated to the 'diagnostic index' in a certain situation gradually increases from zero to a maximum which occurs a little above the discrete threshold value. In this way, a differential spectrum of abnormality may be obtained.

5.3 DISCRETE THRESHOLDS:

The Present Use of Score Functions.

Diagnostic interpretation in the Glasgow program is built on a set of rule-based criteria. These criteria make some use of age, sex, clinical classification, drug therapy and race of the patient but the main criteria are based on whether or not particular ECG measurements attain discrete thresholds. A points scoring system is based on how far an ECG measurement lies above a discrete threshold. These scores can then be added up to give one total index for diagnostic decision making.

In mathematical terms the discrete threshold and subsequent allocation of points can be expressed in terms of a score function $S(x)$ for a measurement x of an ECG variable of interest with a score of K allocated as follows:

$$S(x) = \begin{cases} K & \text{if } x \geq b \\ 0 & \text{if } x < b \end{cases}$$

so that no matter how close to the threshold value b a measurement x is, K points are allocated if it is greater or equal to b and no points are allocated if it is less than b. Figure 5.1 illustrates this discrete score function diagrammatically.

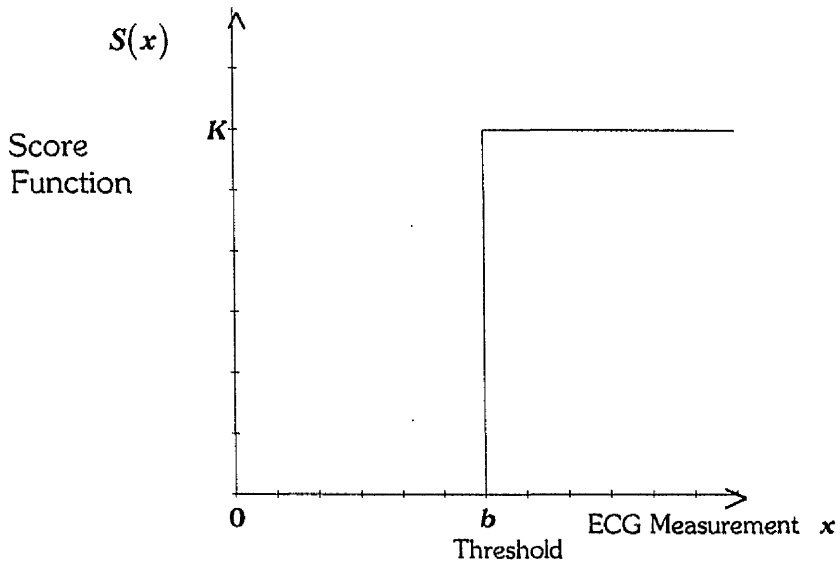


Figure 5.1 **Discrete Score Function (Single threshold)**

An example of the current use of such a score function is in the diagnosis of Left Ventricular Hypertrophy when 2 points are scored if the Lewis Index (which is a criterion based on the sum of the amplitudes of the R wave in lead I and the S wave in lead III minus the sum of the R wave amplitudes in lead I and the S wave in lead III) is greater than a previously defined threshold value, i.e. if

(RIII+SI) - (RI+SIII) > age and sex-dependent limit.

Another contributory factor to the final diagnosis of LVH is evidence of a high R wave amplitude in either lead V5 or V6. Again, 2 points are allocated if the measurement is greater than a specified threshold value and additional increments of 1 are added for every 0.5mV that the voltage exceeds the limit. Figure 5.2 shows the general situation for a voltage measurement x scoring K points when the threshold value b is reached with additional points of 1 being added if x exceeds b by fixed increments c .

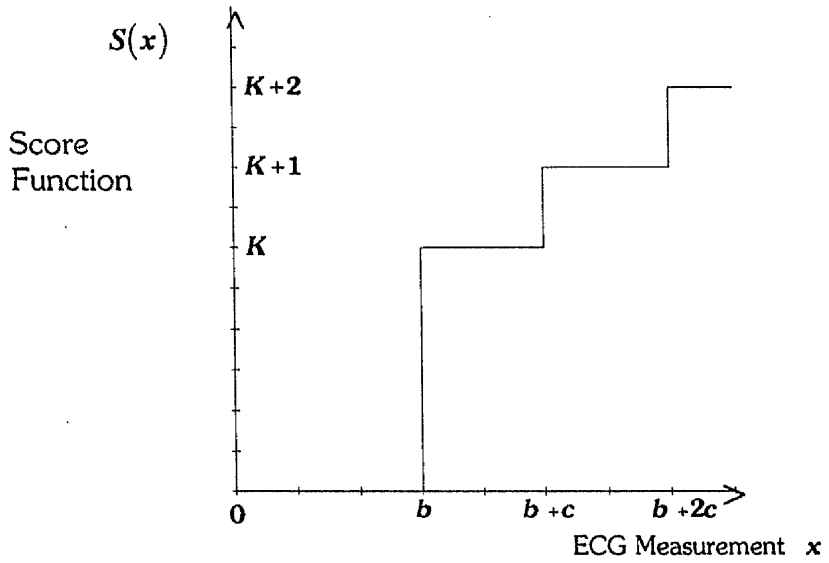


Figure 5.2 **Discrete Score Function (Multiple thresholds)**

5.4 CONTINUOUS THRESHOLDS:

Smoothing a Score Function.

To overcome this discrete threshold problem we will replace the above form of score function with a smoothed version. One appealing form of this is a continuous function for every step. Each step could be replaced by a suitably scaled multiple of a cumulative distribution function (cdf) chosen simply for convenience since any cdf ranges from 0 to 1. We can use the notion of a mean to centre the new function at or close to the old threshold value while the use of a standard deviation, based in our case on the estimate of the day-to-day variability, allows us to dictate the steepness of the new 'smoothed' step.

In all our illustrations we have chosen to use the family of cumulative distribution functions of the logistic distribution simply out of numerical convenience but all of the work could be repeated using any family of cdfs. It would be very surprising if any substantial

differences in the results from our techniques arose from a different choice of underlying family of cdfs. No other version that we have tried has proved to have any substantial effect on our score functions.

The methods which are about to be described should not be interpreted from a probabilistic viewpoint since they are, in the first order, devices based on practical convenience to smooth out discrete boundaries, and we only exploit any probability properties en route to developing smoothed diagnostic indices.

In the single threshold case, a score will be calculated which will depend smoothly on the magnitude of the variable of interest x, increasing from zero to K as x increases. To assess how sharply the 'score' will increase in the vicinity of the thresholds, estimates of the repeat variability as described in 4.4 will be used to dictate an appropriate 'standard deviation' for the probability distribution which is being used as the smooth function. In this case the repeat variation will be represented in the form of day-to-day variability since data is available to estimate this. Incorporating an estimate of the day-to-day variability in the function allows a score to be calculated which takes account of random fluctuation in the measurement x. Thus the bulk of each point will only be allocated if we are sure that the true value x is beyond the old threshold value (accounting for any error).

In this way, an observed value of x which lies some way from the critical value b, but is from an ECG measurement associated with quite a substantial standard deviation σ_x, may be allocated a significant proportion of the score K. In another case it might be that value of x which lies numerically a little nearer to b but with much smaller recording-to-recording variation may score less.

Our chosen smooth score function therefore has the following form :

$$S^{*}{}_{m}(x) = K \times F\{x, \sigma_x; b - \sigma_x\}$$

$$= K \frac{e^{a_x}}{1+e^{a_x}}$$

where $a_x = \dfrac{x-(b-\sigma_x)}{\sigma_x}$.

K = maximum no. of points to be allocated,

b = old threshold value

x = observed value of the ECG variable of interest

σ_x = estimate of day-to-day variability associated with the observation x.

In a somewhat arbitrary manner the distribution has been 'centred' at $(b-\sigma_x)$ rather than at b in order to give a score of $0.7K$ at the old threshold value b. The value $(b-\sigma_x)$ represents a level which is one standard deviation below the old threshold value and ensures that observations falling less than one standard deviation below b will all receive contributions of at least $0.5K$ to the diagnostic index which we feel is a reasonable criterion to adopt.

Figure 5.3 shows the smoothed score function $S_m^{*}(x)$ for the single threshold case while Figure 5.4 demonstrates how this can be extended to the multiple threshold situation simply by summing a series of cdfs, i.e.

$$S_m^{*}(x) = K\left[\frac{e^{a_{1x}}}{1+e^{a_{1x}}} + \frac{e^{a_{2x}}}{1+e^{a_{2x}}} + \frac{e^{a_{3x}}}{1+e^{a_{3x}}} + \ldots\right]$$

where $a_{i_x} = \dfrac{x-(b_i-\sigma_x)}{\sigma_x}$.

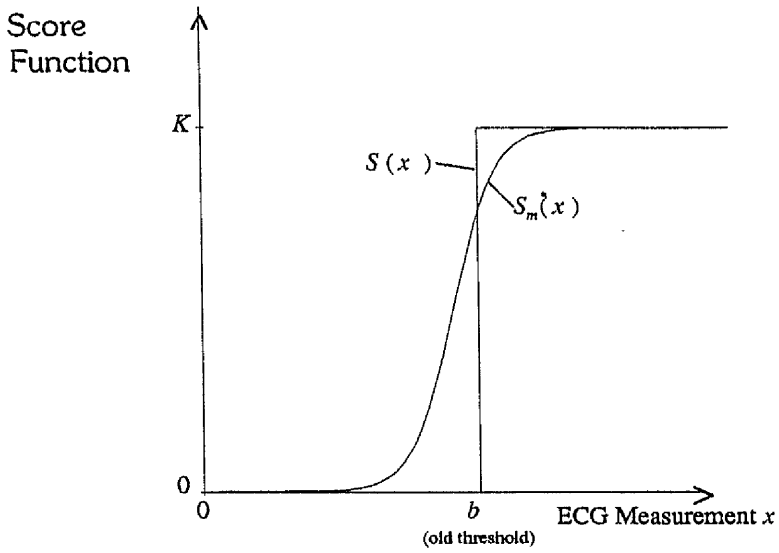


Figure 5.3 **The smoothed score function (Single threshold)**

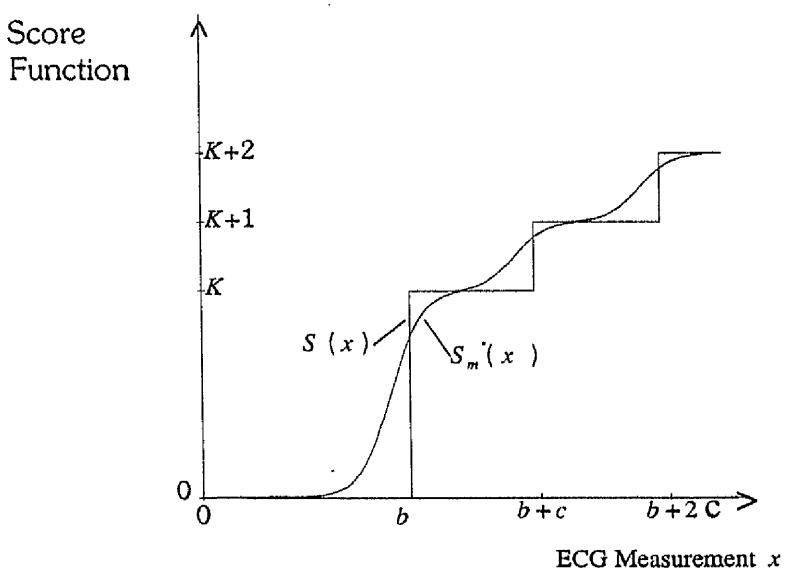


Figure 5.4 **The smoothed score function (Multiple thresholds)**

As a specific example, Figure 5.5 illustrates the smoothed score function specific to the portion of the final LV score associated with the Lewis Index, taking into account the estimated day-to-day variability. This score function will replace the 'old' discrete rule (also represented in Figure 5.5) which was:

$$S(x) = \begin{cases} 2 & \text{if Lewis Index } > \text{age and sex-dependent limit} \\ 0 & \text{otherwise} \end{cases}$$

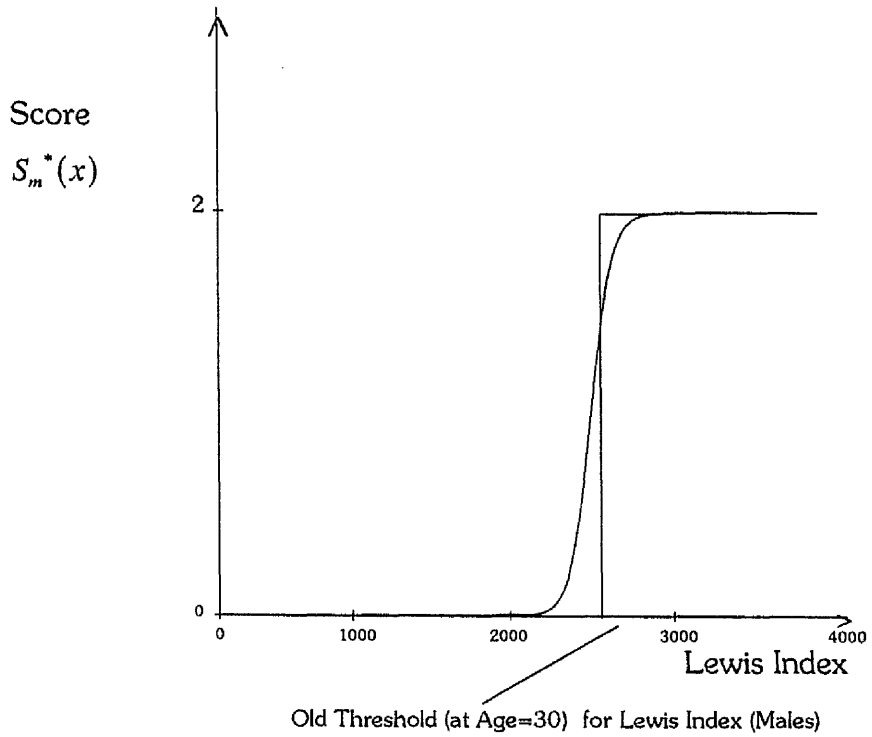


Figure 5.5 **Smoothed score function for Lewis Index**

Similarly, the 'old' discrete rule for RV5 or RV6 which works as follows:

$$S(x) = \begin{cases} 2 & \text{if RV5 or RV6 } > \text{ age and sex-dependent limit} \\ 3 & \text{if RV5 or RV6 } > \text{ limit } + 0.5 \text{ mV} \\ 4 & \text{if RV5 or RV6 } > \text{ limit } + 1.0 \text{ mV} \\ 5 & \text{if RV5 or RV6 } > \text{ limit } + 1.5 \text{ mV} \\ 0 & \text{otherwise} \end{cases}$$

has been replaced by a suitably smoothed version which takes account of the multiple steps. A comparison of the two approaches can be seen in Figure 5.6.

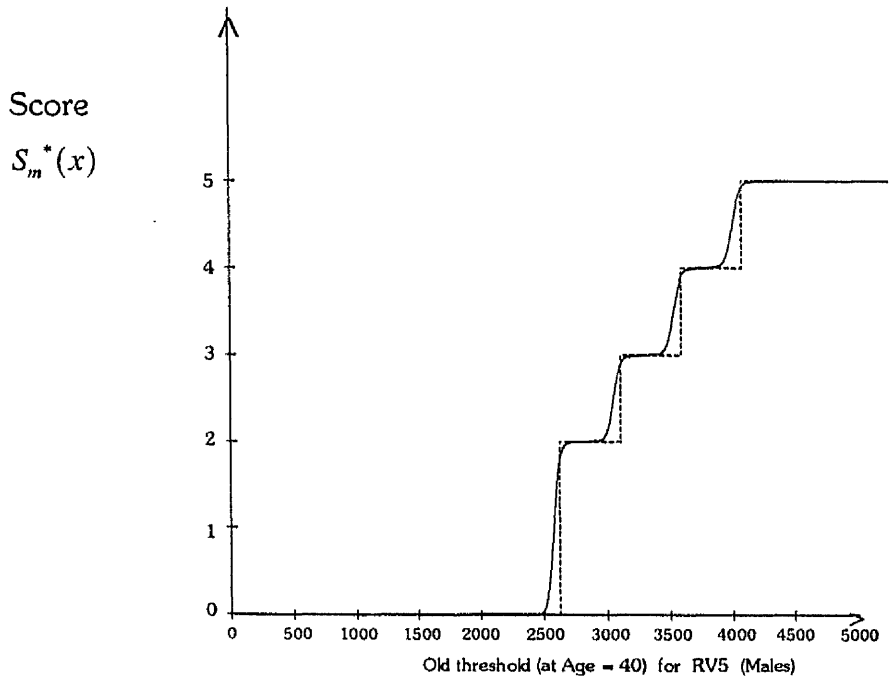


Figure 5.6 Smoothed score function for RV5

5.5 RULES FOR A SMOOTHED COMBINATION OF DIAGNOSTIC CRITERIA

5.5.1 Introduction.

Often, the diagnostic decision is based on a collection of criteria. For example, the deterministic ECG diagnosis of a particular condition may depend on a logical structure involving the union or intersection of various criteria or on a process which relies on the maximum or minimum of several ECG variables being above a specified threshold.

The previously described method which 'smooths over' discrete thresholds in the one-variable case can be used to cater for more complex logical structures by invoking a set of logical combination rules which will now be established.

5.5.2 Two Criteria to be met : The Intersection Rule.

One contributory factor to a final diagnosis of Left Ventricular Hypertrophy in the Glasgow Program is the presence of a negative P terminal force in lead V1. Identification of this condition relies on two separate criteria being satisfied. Two points are scored if this is so while zero points are scored if neither or only one criterion is met.

e.g.

$$S(x_1, x_2) = \begin{cases} 2 & \text{if } x_1 < -100\mu\text{V and } x_2 > 0.04\text{secs} \\ 0 & \text{otherwise} \end{cases}$$

where x_1 = negative amplitude of the negative P wave component in lead V1 and

x_2 = duration of the negative P wave component in lead V1.

The Intersection rule arises as a parallel to a fundamental result of probability theory and may be applied if using the conjunction 'and' when examining two or more criteria. In the general situation, consider variable x_i which is currently associated with threshold value b_i $(i = 1,2)$. The original discrete approach was to assign K points if x_1 is greater than b_1 ***and*** x_2 is greater than b_2, i.e.

$$S(x_1, x_2) = \begin{cases} K & \text{if } x_1 > b_1 \text{ and } x_2 > b_2 \\ 0 & \text{otherwise} \end{cases}$$

Diagrammatically, it can be seen that minute measurement changes in either x_1 or x_2 could result in a major change in the score $S(x_1, x_2)$ from 0 to K or vice versa.

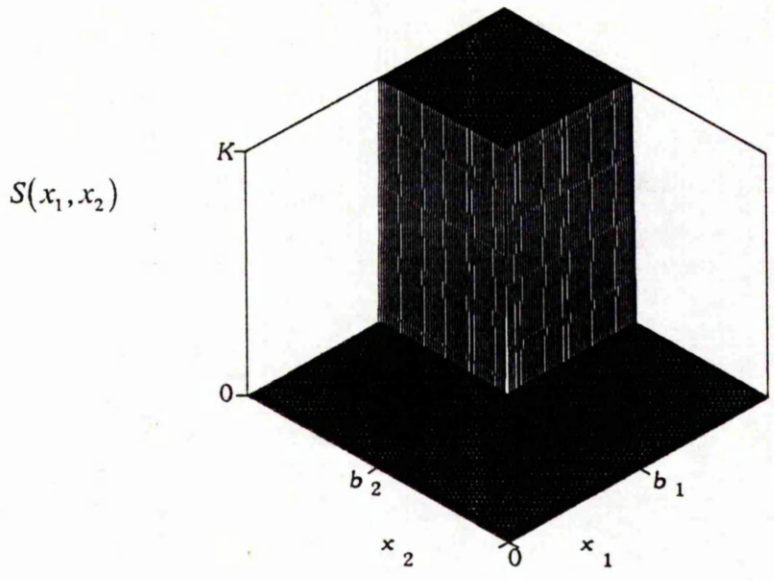


Figure 5.7 **Diagrammatic representation of the Intersection Rule**

A smoothed representation of the score may be based on the one-variable case (see 5.4) with $S_m^*(x_1, x_2)$ increasing from 0 to K as both x_1 and x_2 increase simultaneously.

Using basic rules of probability as an analogue, the smooth score function can be calculated separately for each of the two variables. The overall score is then taken to be the multiple of each contribution, i.e. in the probabilistic analogue

$$\Pr[A_1 \cap A_2] = \Pr(A_1) \times \Pr(A_2)$$

where $A_i = \Pr(X_i > b_i)$. We are effectively adopting the convention of treating the variables x_1 and x_2 as statistically independent although in practice this is almost certainly untrue. The rationalisation for such a move is simple practical convenience.

Translating this principle to the situation described, we can write

$$S_m^*(x_1, x_2) = K \times F_1 \times F_2$$

where $F_i = F\{x_i, \sigma_{x_i}; b_i - \sigma_{x_i}\} = \dfrac{e^{a_{x_i}}}{1+e^{a_{x_i}}}$, $a_{x_i} = \dfrac{x_i - (b_i - \sigma_{x_i})}{\sigma_{x_i}}$, x_i is the observed value of the ECG variable of interest, b_i the threshold value

and σ_{x_i} the estimated day-to-day variability associated with the observation x_i .

As before, the distribution has been 'centred' at $\left(b_i - \sigma_{x_i}\right)$ to ensure individual scores of approximately $0.7K$ when each of the discrete threshold values are met. Figure 5.9 illustrates the smooth representation of the Intersection Rule.

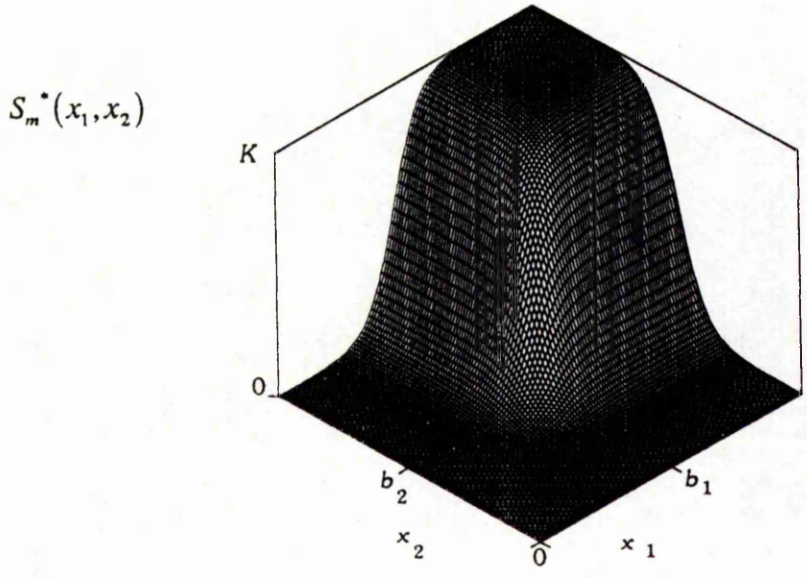


Figure 5.8 **Smooth representation of the Intersection Rule**

5.5.2 One or other Criterion to be met: The Union Rule.

The second basic combination rule occurs when a contribution to the score is made if ***either*** one ***or*** other of two threshold criteria are met. An example of this can be found once again in the diagnosis of LVH when one point is added to the LVH diagnostic index if the intrinsicoid deflection in either of leads V5 or V6 is greater than 60 msecs, i.e.

$$S(x_1, x_2) = \begin{cases} 1 & \text{if } x_1 > 60 \text{ msecs or} \\ & \quad x_2 > 60 \text{ msecs} \\ 0 & \text{otherwise} \end{cases}$$

In the general situation the Union Rule is applied when considering a combination of two criteria, although this time interest is in the use of the conjunction 'or' . Consider the situation where K points are to be allocated if x_1 is greater than b_1 **or** x_2 is greater than b_2, i.e.

$$S(x_1, x_2) = \begin{cases} K & \text{if } x_1 > b_1 \text{ or } x_2 > b_2 \\ 0 & \text{otherwise} \end{cases}$$

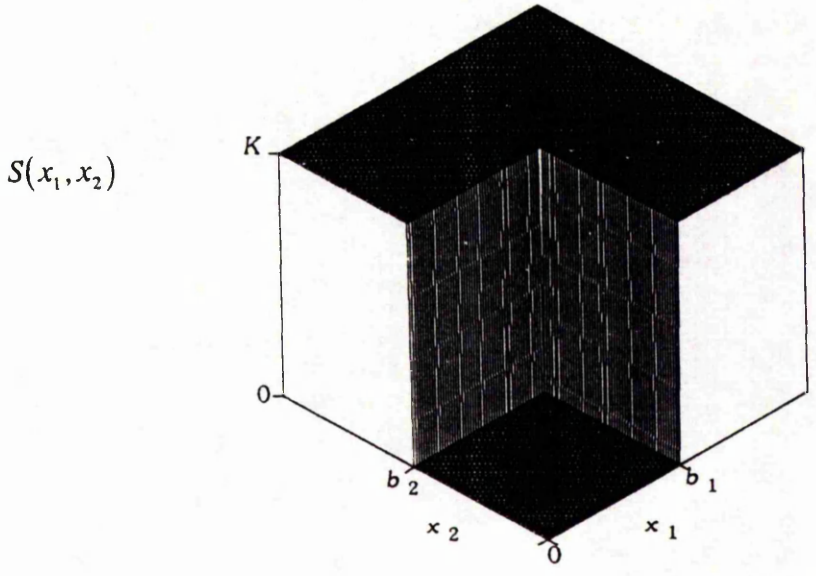


Figure 5.9 **Diagrammatic representation of the Union Rule**

Again, differences in the score may occur if there is a small measurement change in either x_1 or x_2 .

The following simple rule of probability theory may be used as an analogue to this situation:

$$\Pr[A_1 \cup A_2] = \Pr(A_1) + \Pr(A_2) - \Pr(A_1 \cap A_2)$$
$$= \Pr(A_1) + \Pr(A_2) - \Pr(A_1) \times \Pr(A_2)$$

again exploiting an 'assumption' of independence. When this is applied to the diagnostic criteria context we have:

$$S_m^*(x_1, x_2) = K \times [F_1 + F_2 - F_1 \times F_2]$$

where F_i are defined as before. Figure 5.11 shows the smooth representation of the Union Rule.

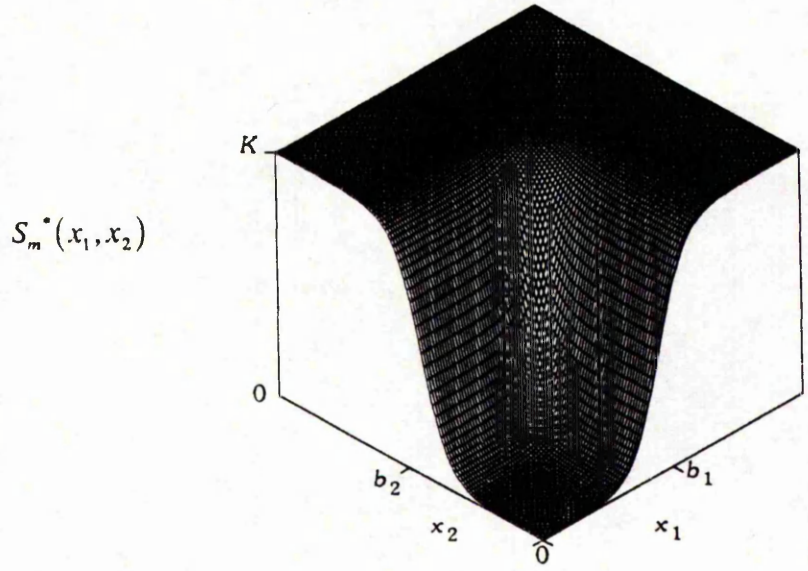


Figure 5.10 Smooth representation of the Union Rule

5.5.4 More than two Criteria to be met:

A) The Minimum Rule.

Certain stages of the deterministic process of ECG diagnosis rely on the maximum or minimum of a number of variables achieving a specified threshold.

This approach can also be treated by extending the basic ideas used in the union and intersection rules in order to smooth out any discrete thresholds encountered.

For example, if all of a certain combination of ECG variables must be beyond a specified threshold value, i.e.

$$S(x) = \begin{cases} K & \text{if } \min(x_1, x_2, x_3) > b \\ 0 & \text{otherwise} \end{cases}$$

then we can write

$$S_m^*(x_1, x_2, x_3) = \begin{cases} K & \text{if all } (x_1, x_2, x_3) > b \\ 0 & \text{otherwise} \end{cases}$$

$$= \begin{cases} K & \text{if } x_1 > b \text{ and } x_2 > b \text{ and } x_3 > b \\ 0 & \text{otherwise} \end{cases}$$

Applying the same argument as for the intersection rule it is natural to write this as

$$S_m^*(x_1, x_2, x_3) = K \times F_1 \times F_2 \times F_3$$

where $F_i = \dfrac{e^{a_{x_i}}}{1+e^{a_{x_i}}}$; $a_{x_i} = \dfrac{x_i - (b - \sigma_{x_i})}{\sigma_{x_i}}$.

This can obviously be extended to any number of variables.

B) The Maximum Rule.

If the overall criterion is that at least one of a certain combination of ECG variables has to be greater than a specified threshold then we have

$$S(x_1, x_2, x_3) = \begin{cases} K & \text{if } \max(x_1, x_2, x_3) > b \\ 0 & \text{otherwise} \end{cases}$$

Consider the equivalent probability result

$$\Pr[\max(x_1, x_2, x_3) > b]$$

which can be rewritten as

$$1 - \Pr[\max(x_1, x_2, x_3) < b]$$

By the previous 'independence' analogy, it seems reasonable to write this as

$$1-\Pr\left[\max(x_1,x_2,x_3)<b\right] = 1-\Pr(x_1<b)\Pr(x_2<b)\Pr(x_3<b)$$
$$= 1-(1-F_1)\times(1-F_2)\times(1-F_3)$$

where F_i is in the same form of the previous section.

Hence the equivalent formulation of the score function is

$$S_m^*(x_1,x_2,x_3) = K\left[1-(1-F_1)\times(1-F_2)\times(1-F_3)\right]$$

where clearly for 2 variables this reduces to the form used in section 5.5.2. Again, we can extend the idea to any number of variables.

5.5.5 Combinations of Criteria.

Since most diagnostic criteria involve combinations of unions, intersections, maxima and minima we can now handle these by using the basic logical building blocks developed already in this chapter. We now illustrate this process through several examples.

EXAMPLE 1

Consider the score function

$$S(x_1,x_2,x_3,x_4) = \begin{cases} K & \text{if } \max(x_1,x_2)>b_1 \text{ and } \min(x_3,x_4)>b_2 \\ 0 & \text{otherwise} \end{cases}$$

which can be written in a smoothed form as

$$S_m^*(x_1,x_2,x_3,x_4) = K\times\left[1-(1-F_{11})\times(1-F_{21})\right]\times F_{32}\times F_{42}$$

where $F_{ij}=\dfrac{e^{a_{x_{ij}}}}{1+e^{a_{x_{ij}}}}$ and $a_{x_{ij}}=\dfrac{x_i-(b_j-\sigma_{x_i})}{\sigma_{x_i}}$.

EXAMPLE 2

Here, consider the score function

$$S(x_1, x_2, x_3, x_4) = \begin{cases} K & \text{if } \min(x_1, x_2) > b_1 \text{ or } \max(x_3, x_4) > b_2 \\ 0 & \text{otherwise} \end{cases}$$

which can be written in the smoothed form as

$$S_m^*(x_1, x_2, x_3, x_4) = K \times \left[F_{11}F_{12} + (1 - F_{32}) \times (1 - F_{42}) - F_{11}F_{12} \times (1 - F_{32}) \times (1 - F_{42}) \right]$$

where the F_{ij} are in the same form as in Example 1.

Having organised the logical structure of smoothing combinations of discrete threshold-based criteria, we now illustrate its effects on the diagnosis of ST-T changes.

5.6 AN ILLUSTRATION OF A SMOOTHED DIAGNOSTIC INDEX: ST-T Changes.

One area of the diagnostic program which assesses the severity of a particular condition is in the identification of ST-T changes.

This area is notorious as one of the most difficult aspects of electrocardiography not least because the line dividing normal from abnormal is not sharp and normal ranges for components of the ST segment and T wave are wide (Macfarlane and Lawrie, 1989d). Clinically, there may be many possible causes of changes in the ST-T segment which further complicates any diagnosis.

This area will be discussed in a little more detail in Chapter 7 although reference will now be made to the interpretation of ST

depression and ST elevation since this is accomplished on the basis of the magnitude of two variables, namely ST amplitude and ST slope. Abnormalities of the ST segment are defined on the basis of the ST amplitude and slope, a clearly negative segment indicating marked ST depression and an obviously positive segment indicating significant ST elevation. Both ST depression and ST elevation are rather arbitrarily defined as moderate, marked or severe depending on the amplitude and slope, thereby providing a spectrum of the response. In this situation the response takes the form of a score which ranges from -3 to +3 describing the transition from marked ST depression through to significant ST elevation, i.e. there is a score function $S(x_1, x_2)$ defined by the rule:-

if	$x_1 < -100\mu V$	and $x_2 < 0°$	then	$S(x_1, x_2) = -3$
	$x_1 < -50\mu V$	and $x_2 < 0°$	then	$S(x_1, x_2) = -2$
	$x_1 < -20\,\mu V$	and $x_2 < 0°$	then	$S(x_1, x_2) = -1$
	$x_1 > 60\mu V$	and $x_2 > 0°$	then	$S(x_1, x_2) = 1$
	$x_1 > 80\mu V$	and $x_2 > 0°$	then	$S(x_1, x_2) = 2$
	$x_1 > 100\mu V$		then	$S(x_1, x_2) = 3$
	otherwise			$S(x_1, x_2) = 0$

where x_1 = ST Amplitude and x_2 = ST slope. This is illustrated diagrammatically in Figure 5.11.

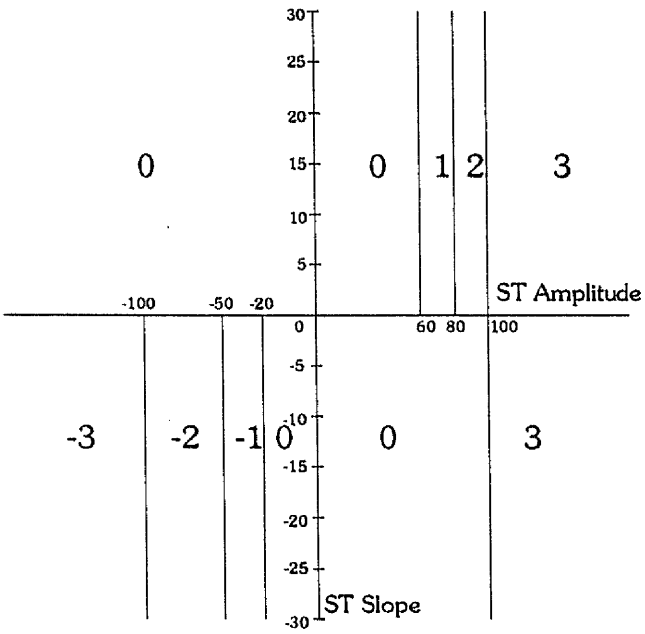


Figure 5.11 **Contour plot of the discrete ST Score $S(x_1, x_2)$**

These criteria apply to lateral and inferior leads (I, II, III, aVL, aVF, V5 and V6) only. Slightly different threshold values for the amplitudes are used when considering leads V1, V2, V3 and V4.

Clearly there is potential for changes in the score due to small measurement changes especially the sudden jump from 0 to 3 in the lower right quadrant (see Figure 5.11). The methods which have been described earlier in this chapter can be used to form a smooth representation of $S(x_1, x_2)$, namely $S^*_m(x_1, x_2)$.

The negative contribution to $S^*_m(x_1, x_2)$ is calculated on the basis of the magnitude of the negative ST amplitude and the slope of the ST segment in the appropriate lead. This part of the score can range from 0 to -3 and may be expressed as follows :

$$S^*_{m_1}(x_1, x_2) = -1 \times \{(1 - F_{11}) + (1 - F_{12}) + (1 - F_{13})\} \times (1 - F_{21})$$

The positive contribution can be represented in a similar way.

$$S^*_{m_2}(x_1, x_2) = \{F_{14} + F_{15} + F_{16}\} \times F_{21}$$

where x_1 = ST Amplitude in lead of interest;

σ_{x_1} = estimated day - to - day variability in ST Ampl.;

x_2 = ST Slope in lead of interest;

σ_{x_2} = estimated day - to - day variability in ST Slope.

$$F_{1i} = \frac{e^{b_{i_{x_1}}}}{1+e^{b_{i_{x_1}}}};\quad b_{i_{x_1}} = \frac{x_1 - (b_i - \sigma_{x_1})}{\sigma_{x_1}};\quad F_{21} = \frac{e^{a_{1_{x_2}}}}{1+e^{a_{1_{x_2}}}};\quad a_{1_{x_2}} = \frac{x_2 - (a_1 - \sigma_{x_2})}{\sigma_{x_2}}$$

$b_1 = -100;\quad b_2 = -50;\quad b_3 = -20;$

$b_4 = 60;\quad b_5 = 80;\quad b_6 = 100;$

$a_1 = 0.$

The overall smoothed score can now be calculated as the sum of the separate components $S^*_{m_1}(x_1, x_2)$ and $S^*_{m_2}(x_1, x_2)$ since neither of these will contribute substantially to the other in the appropriate ranges, i.e.

$$S^*_m(x_1, x_2) = S^*_{m_1}(x_1, x_2) + S^*_{m_2}(x_1, x_2)$$

Figures 5.12 and 5.13 illustrate the surface plots of the score functions based on the discrete and on the new smoothed methods respectively. Figure 5.13 demonstrates the advantage that $S^*_m(x_1, x_2)$ has over $S(x_1, x_2)$ (see Figure 5.12) in terms of continuity.

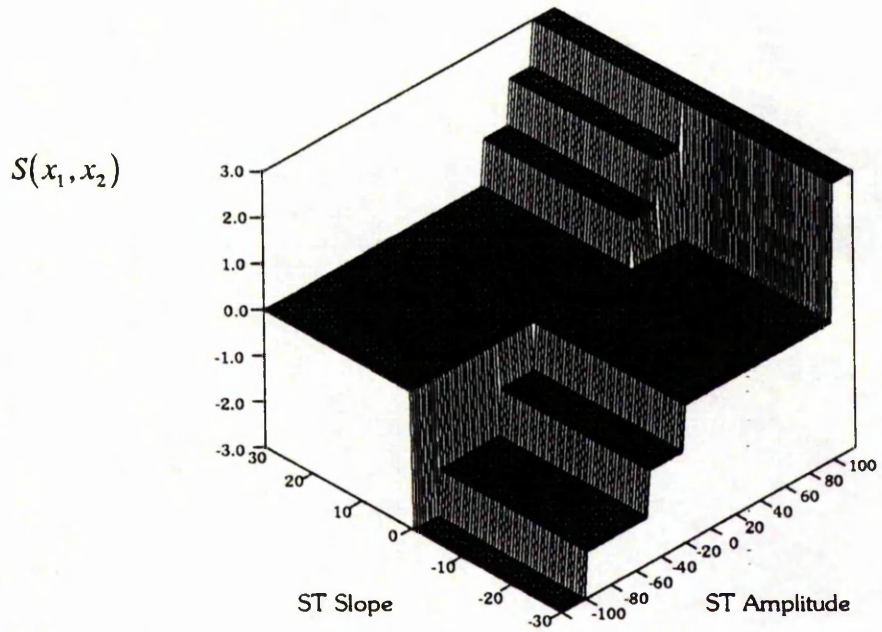


Figure 5.12 **Surface plot of $S(x_1, x_2)$**

$S_m^*(x_1, x_2)$

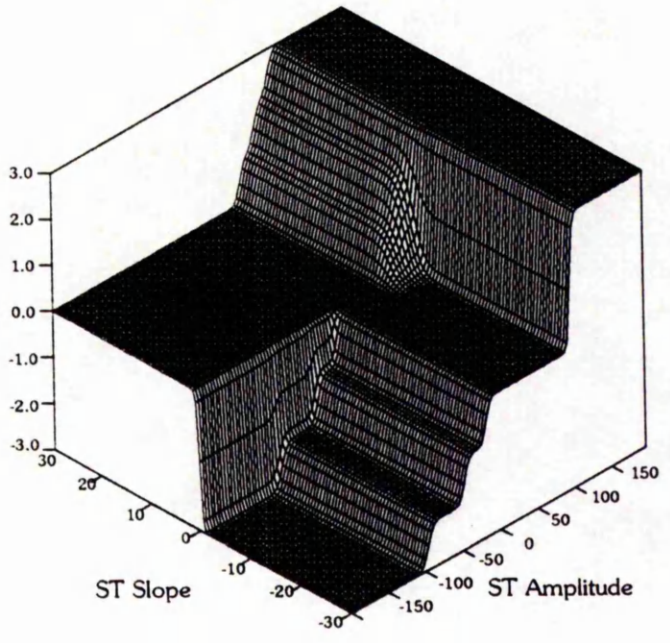


Figure 5.13 **Surface plot of $S_m^*(x_1, x_2)$**

5.7 SUMMARY.

This chapter has introduced the concept of 'smoothing out' discrete thresholds by replacing them with continuous functions. These functions, which are based on cumulative distribution functions for computational convenience rather than on any principle of probability theory, are used to provide smoothed versions of previously discrete diagnostic indices in both the single and multiple threshold situations. They have an additional advantage of taking into account the natural day-to-day variability occurring in each ECG measurement which dictates the amount of 'steepness' associated with each new smoothed diagnostic index.

Since many diagnostic decisions are based on collections of combined criteria we have also evolved a methodology for the

treatment of such combinations. The union rule can be used to smooth out the diagnostic index in the situation where a certain action is to be taken when ***either*** one of two conditions is met. Similarly, the intersection rule may be implemented if ***both*** criteria are to be satisfied simultaneously before an action is taken. Combinations of the ***maxima*** and ***minima*** rules may be manipulated to cater for situations when more than two conditions are being considered.

Basing an algebra of score construction on these criteria allows us to smooth out many complicated forms of 'discrete' rules which are encountered throughout the diagnostic process.

CHAPTER SIX:

LEFT VENTRICULAR HYPERTROPHY.

6.1 INTRODUCTION

In this Chapter, the term hypertrophy will be used to cover increased ventricular wall thickness and/or increased volume. Ventricular hypertrophy arises as a result of pressure or volume overload on the heart and may produce significant changes in the ECG. Each ventricle may be affected independently of the other, causing either left or right ventricular hypertrophy. Biventricular hypertrophy arises if both ventricles are enlarged simultaneously.

6.2 LEFT VENTRICULAR HYPERTROPHY - EVOLUTION OF ECG CRITERIA

6.2.1 Voltage Criteria

Left ventricular hypertrophy is normally characterised by increased QRS voltages in the left ventricular leads (I, aVL, V5 and V6) and in V1 and V2.

Evolution of voltage criteria for the detection of Left Ventricular Hypertrophy (LVH) began in the early 20th century when Lewis (1914) established an index of R and S voltages in the limb leads I and III, i.e.

$$(RI + SIII) - (RIII + SI).$$

He observed that an index of 1.7mV or more was indicative of LVH (with a sensitivity of 18% and a specificity of 98%) and later research substantiated this claim (Hermann and Wilson, 1922).

Later, Wilson introduced a voltage criterion suggesting that an S wave amplitude of greater than 2.4mV in lead V1 was consistent with a hypertrophied left ventricle (Wilson, 1944).

Sokolow and Lyon (1949) defined several additional voltage criteria specific to the precordial leads, the most extensively used being the Sokolow-Lyon Index, namely

$$SV1 + RV5 > 3.5mV$$

6.2.2 Non-voltage Criteria

One of the first non-voltage criteria for the diagnosis of ECG LVH was the presence of left axis deviation demonstrated by Gubner and Ungerleider (1943). The cardiac axis represents the average spread of the depolarisation wave through the ventricles and if the left ventricle is hypertrophied this may cause the axis to rotate towards the left.

It was also observed that a delay in the onset of the intrinsicoid deflection in the left ventricular leads may also be indicative of LVH (Noth, Myers and Klein 1947). Intrinsicoid deflection represents the time taken for the electrical impulse to spread through the ventricles to the area beneath the electrode so that any increase in the ventricular mass is likely to prolong this interval.

Any increase in the duration of the QRS complexes in the left ventricular leads may also be used as a contributory criterion because the QRS duration represents ventricular depolarisation which will take longer if the ventricular mass is greater.

6.3 EVOLUTION OF ST CONTOUR CRITERIA

Occasionally, LVH can be accompanied by structural and metabolic changes in the heart muscle which result in distortion of the ST-T segment. Such changes are often seen as depression of the ST segment and inversion of the T wave and although these manifestations may not solely relate to an increase in mass of the left ventricle, when found in the presence of other LVH-related criteria, they suggest a poorer prognosis (Milliken, Macfarlane and Lawrie, 1989). Therefore it is of the utmost importance to be able to diagnose LVH with evidence of ST-T changes (known as LV strain) accurately.

6.4 THE GLASGOW SCORING SYSTEM FOR LVH

Romhilt and Estes developed a system which relied on both voltage and non-voltage criteria for ECG diagnosis of LVH (Romhilt and Estes, 1968). This point scoring system has been incorporated into some computer programs and has been modified for use in the Glasgow laboratory (Huwez, 1991) see Table 6.1.

Criterion	No. of Points
Increased QRS Voltage	≥2 points
ST-T changes	1 - 4 points
Increased P terminal force	2 points
Left Axis Deviation	2 points
Prolonged QRS Duration	1 point
Delayed Intrinsicoid Deflection	1 point

Table 6.1 Brief outline of the Glasgow scoring system for the detection of LVH

The final LVH score is calculated by summing the points recorded by the six individual components and the diagnosis made in the following way:

If LVH Score	≤ 3	then	"No LVH"
	= 4	then	"Possible LVH"
	= 5	then	"Probable LVH"
	≥ 6	then	"Definite LVH".

The presence of discrete threshold values throughout the Glasgow scoring system for LVH has, as previously outlined, contributed to a lack of repeatability between consecutive recordings which may have arisen as a result of small measurement changes in the neighbourhood of these boundaries.

Certain areas of the process for the diagnosis of LVH are somewhat complex. In the main, the smoothing techniques already described have been implemented in order to minimise the effects of small day-to-day and repeat variation on the overall diagnostic criteria.

6.4.1 Smoothing the Voltage Score

In the LVH scoring system, points are allocated if a particular voltage measurement lies above a threshold value and in the Glasgow program these limits of 'healthiness' are age and sex-dependent. Chapter 3 described a method of replacing these discrete upper limits of normal with continuous equations and this now forms the basis of the new approach. The level of day-to-day variability in each voltage has also been estimated and used towards providing stability in the construction of the new 'smooth' LVH score.

An example of the new method for calculating the portion of the LVH score $S_m^*(x)$ attributable to any increase in QRS voltage is given, in this case the contribution made by the RV5 voltage for Caucasian males aged 18 years and upwards.

$$S^*_m(x) = 2\times\left[F\{x,\sigma_x;b-\sigma_x\}+F\{x,\sigma_x;(b+0.5)-\sigma_x\}+\right.$$
$$\left.F\{x,\sigma_x;(b+1.0)-\sigma_x\}+F\{x,\sigma_x;(b+1.5)-\sigma_x\}\right]$$

using the notation introduced in Chapter 5 (section 5.4). The value x denotes the R amplitude in lead V5, b is the continuous equation describing the upper limit of 'healthiness' for this measurement and σ_x^2 is the estimated amount of day-to-day variability in measurement x, i.e.

$$b = (59.77 - 0.01089\times \text{Age})^2;$$
$$x = \text{RV5};$$
$$\sigma_x = \sqrt{(32^2 + 3.72^2 \times x)}.$$

A similar smoothed version of this part of the scoring function may be calculated for female Caucasians and for other races by replacing b by the relevant upper limit of 'healthiness'. A comparison

of the old and new methods of assigning scores to the magnitude of the R wave in lead V5 for male Caucasians of 40 years of age can be seen in Figure 5.6.

Identification of a negative P terminal force in lead V1 (the third of the criteria listed in Table 6.1) and its subsequent contribution to the LVH score relies on two criteria being satisfied simultaneously. Thus the techniques described in Chapter 5 can be used to smooth out the two discrete boundaries.

An abnormal P terminal force in lead V1 is recognised if the amplitude of the P wave is more negative than -110 μV and the duration is greater than 40 msecs. Since 2 points are allocated if such a combination of criteria is met (see Table 6.1), the discrete approach can be described as follows:

$$S(x_1, x_2) = \begin{cases} 2 & \text{if } x_1 < -110\mu V \text{ and} \\ & \quad x_2 > 40 \text{ msecs} \\ 0 & \text{otherwise} \end{cases}$$

where x_1 = negative of the amplitude of the negative P wave component in lead V1

and x_2 = duration of the negative P wave component in lead V1.

This is illustrated diagramatically in Figure 6.1.

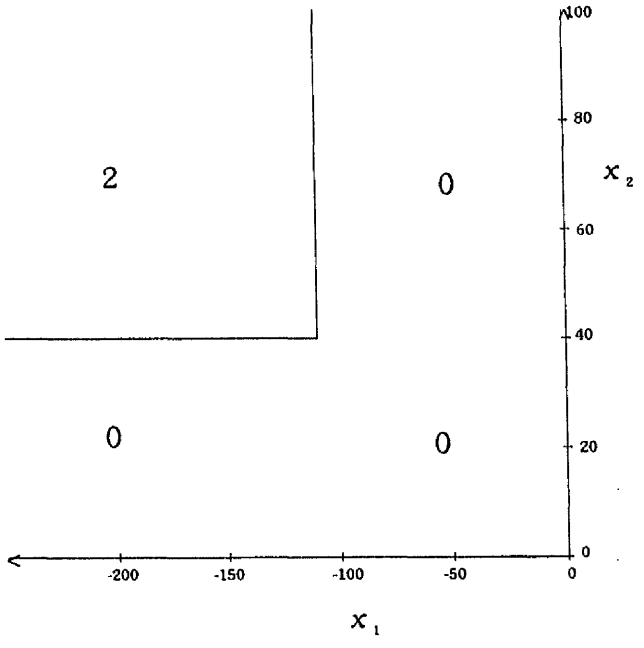


Figure 6.1 **Diagrammatic representation of the discrete score function associated with P terminal force in lead V1.**

Using the smoothing techniques, this becomes

$$S_m^*(x_1, x_2) = 2 \times \left(1 - F\left\{x_1, \sigma_{x_1}; b_{x_1} - \sigma_{x_1}\right\}\right) \times F\left\{x_2, \sigma_{x_2}; b_{x_2} - \sigma_{x_2}\right\}$$

where $b_{x_1} = -110$; $x_1 = -\mathrm{P}_{\mathrm{neg}}\mathrm{Amp.\ V_1}$;

$$\sigma_{x_1} = \sqrt{\left(0.60^2 + 0.18^2 \times |x_1|\right)} \ ;$$

$b_{x_2} = 40$; $\quad x_2 = \mathrm{P}_{\mathrm{neg}}\mathrm{Durn.\ V_1}$;

$$\sigma_{x_2} = \sqrt{\left(10.8^2 + 0.08^2 \times x_2\right)}$$

and this new smooth form of scoring can be seen in Figure 6.2.

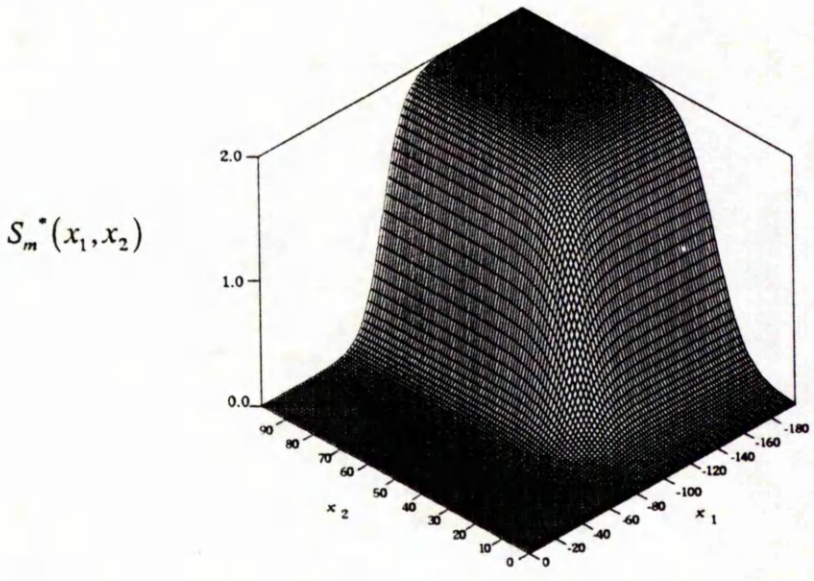


Figure 6.2 **Diagrammatic representation of the smooth score function associated with P terminal force in lead V1.**

6.4.2 Smoothing the Non-voltage Score

As mentioned previously, any prolongation in the Intrinsicoid Deflection in the left ventricular leads can be suggestive of LVH and the current program adopts the following strategy for contributing points to the final LVH score (see Figure 6.3).

$$S(x_1, x_2) = \begin{cases} 1 & \text{if } x_1 > 60 \text{ msecs or } x_2 > 60 \text{ msecs} \\ 0 & \text{otherwise} \end{cases}$$

where x_1 = intrinsicoid deflection in lead V5 and

x_2 = intrinsicoid deflection in lead V6.

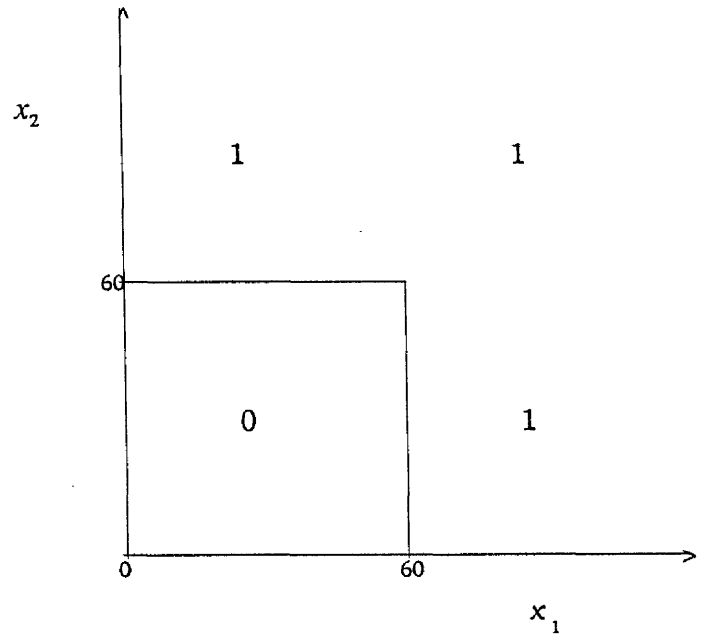


Figure 6.3 Diagrammatic Representation of the discrete score function associated with Intrinsicoid Deflection.

The equivalent smooth version (which is illustrated in Figure 6.4) is

$$S_m^*(x_1,x_2) = F\{x_1,\sigma_{x_1};b-\sigma_{x_1}\}+F\{x_2,\sigma_{x_2};b-\sigma_{x_2}\}- F\{x_1,\sigma_{x_1};b-\sigma_{x_1}\}\times F\{x_2,\sigma_{x_2};b-\sigma_{x_2}\}$$

where b = 60 msecs, x_1 = Int. Deflection in V5 (msecs), x_2 = Int. Deflection in V6 (msecs) and ${\sigma_{x_1}}^2, {\sigma_{x_2}}^2$ denote the estimated amount of day-to-day variability observed in the Int. Deflection in V5 and V6 respectively, i.e.

$$\sigma_{x_1} = \sqrt{(16.0+0.22^2\times x_1)} \;;\; \sigma_{x_2} = \sqrt{(27.04+0.16^2\times x_2)}.$$

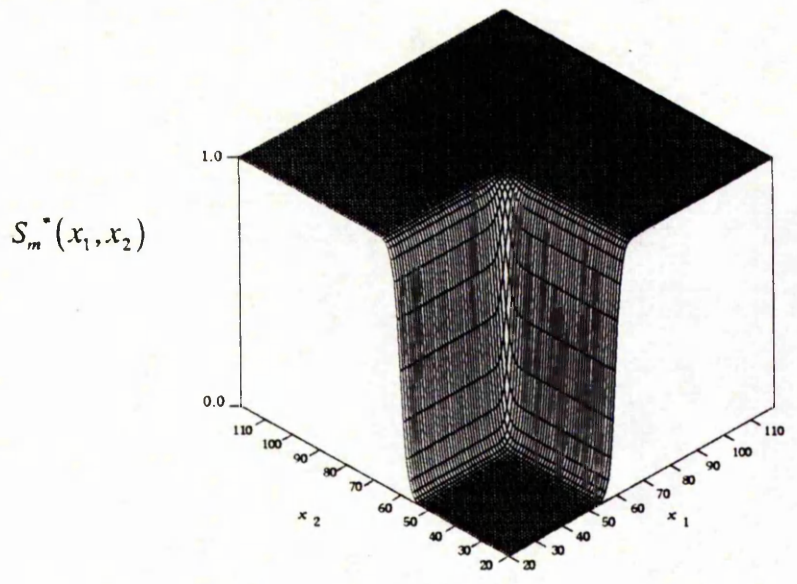


Figure 6.4 **Diagrammatic representation of the smooth score function associated with Intrinsicoid Deflection.**

6.4.3 Smoothing the LV Strain Score

In the Glasgow program the presence of secondary ST-T changes (or left ventricular strain) contributes significantly to the LVH scoring system. The six criteria involved are listed below, 4 points being scored if parts (i) to (vi) all hold and 2 points if all but part (iv) are satisfied. These points are then added to the LVH score.

(i) $ST_j \leq -0.02$mV and ST slope $\leq -5°$ **OR**
$ST_j \leq -0.05$mV and ST slope $\leq 0°$

(ii) $|ST_j - T_{neg}| > 0.10$ mV

(iii) $T_{neg} < 0.20$mV **AND** $T_{morph} < 0$ with $T_{pos} < 0.15$mV

(iv) R (or R') > 1.0mV

(v) No Q waves in lateral leads

(vi) $QRS_{dur} < 120$msecs

where ST_j is the amplitude of the ST segment in the lead j, j = I, aVL, V5 and V6,

T_{pos} and T_{neg} are the amplitudes of the maximum positive and negative parts of the T wave respectively and

T_{morph} is the morphology of the T wave.

Diagnosis of LV Strain as a contribution to LVH also requires a smooth technique which has the ability to cope with the complexities of the outlined criteria. The nature of the criteria relating to the identification of LV Strain is such that small measurement changes in the neighbourhood of threshold values may alter the final diagnosis of secondary ST-T changes and hence LVH. The simplest way to observe this is to assume that parts(i), (ii), (iii), (v) and (vi) hold from one recording to another but that the R amplitude in any of the leads I, aVL, V5 or V6 is 1.0mV on day 1 and 1.1mV on day 2. Two points would be added to the LV score for the first recording and four points for the second. In the worst scenario this difference of two points could result in an individual being diagnosed as normal on one occasion and as having possible LVH on the next.

Assigning a score to each of the six criteria independently will form the basis on which to develop a smoothed representation of LV strain. This may be done arbitrarily as follows:

Score 2 if ($ST_j \leq$ -0.02mV and ST slope $\leq$ -5° **OR**
$ST_j \leq$ -0.05mV and ST slope $\leq$ 0°)
AND $|ST_j - T_{neg}| > 0.10$ mV

Score $1\frac{3}{4}$ if T_{neg}<0.20mV **AND** $T_{morph} < 0$ with
T_{pos}<0.15mV
AND R (or R') > 1.0mV

Score $\frac{1}{4}$ if $QRS_{dur} < 120$msecs

(N.B. No score will be recorded for LV Strain if significant Q waves exist in any of the leads I, aVL, V5 or V6). This will effectively provide a "smoothed version" of parts (i) to (vi) described earlier.

Once again the combination of ST slope and ST amplitude forms the basis of part of the score function and the techniques outlined in Chapter 5 may be used.

Smoothing of the separate stages of the score function is done in the usual way and the results combined using the union and intersection rules which have been described in Chapter 5.

6.5 PERFORMANCE OF THE SMOOTH TECHNIQUE

The new approach described above has been developed with the aim of minimising the effect of day-to-day variation on the diagnostic criteria. It is not expected to enhance the diagnostic accuracy of the existing deterministic program to any great extent but nevertheless it needs to perform comparably with respect to sensitivity and specificity.

It was necessary to write the logic for the calculation of the new smoothed LVH index in Fortran. This proved to be a lengthy procedure which involved extensive restructuring of certain sections of the existing Glasgow program. A small section of the new diagnostic coding can be seen in part A) of the Appendix.

Once accomplished, it was then possible to make a formal comparison between both versions of the Glasgow program. Each version has been used to analyse a database of ECGs comprising 84 clinically documented LVH cases and 136 non-LVH cases.

Figure 6.5 illustrates the separation obtained between LVH and non-LVH patients when the existing program is used. The discrete LVH score is calculated by summing the six separate contributions received which were summarised in Table 6.1.

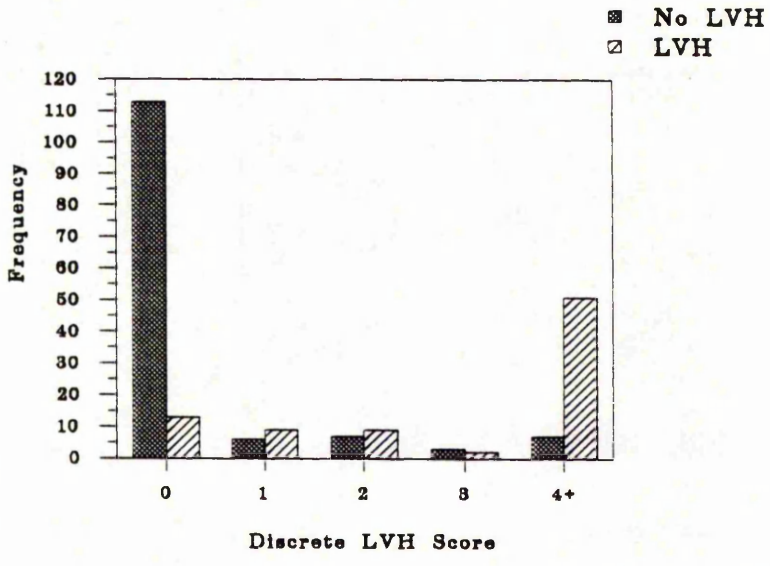


Figure 6.5 **Frequency Distribution of the Discrete LV Score (based on 136 Non LVH and 84 LVH cases) (See Table 6.1 for definition of Discrete Score).**

The scores obtained using the smooth technique are of a continuous nature, thus requiring selection of an optimal cutpoint. Receiver Operating Characteristic (ROC) curves were used to determine the 'best' value in terms of separating 'normals' from LVH cases, the value closest to the origin being chosen as corresponding to the cutpoint which minimises the sum of the number of false positives and the number of false negatives (Macfarlane, 1989d).

The ROC curve produced for different cutpoints of the smooth score suggests that the optimal cutpoint which will produce the best results is approximately 3.5 (see Figure 6.6).

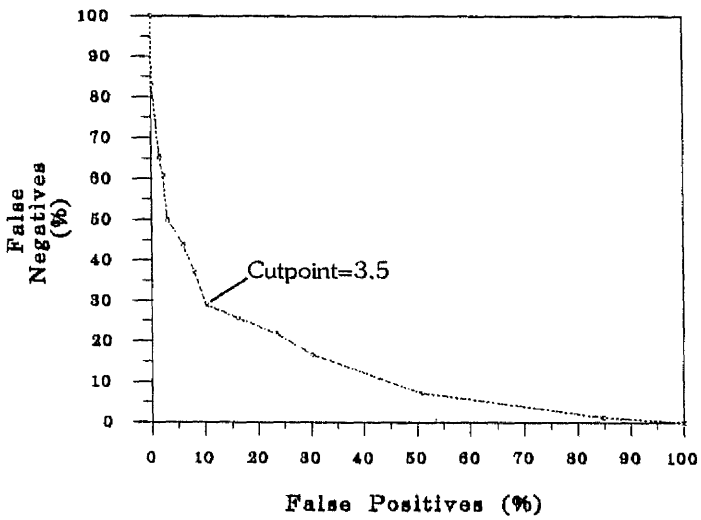
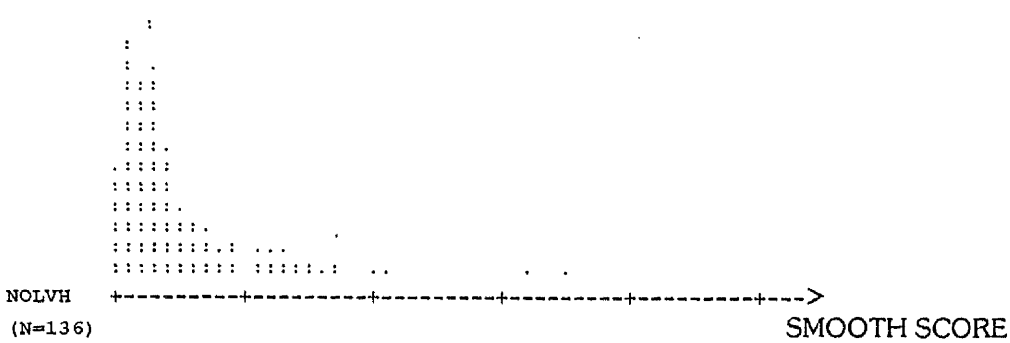


Figure 6.6 **Receiver Operating Characteristic Curve illustrating the optimum cutpoint of 3.5**

Figure 6.7 illustrates the separation of the two groups based on such a cutpoint.

DISCRETE SCORE ≤ 3



DISCRETE SCORE ≥ 4

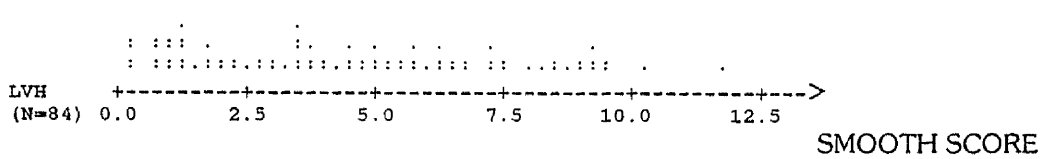


Figure 6.7 **Frequency Distribution of the smoothed LV Score (based on 136 Non LVH and 84 LVH cases)**

This value produces a sensitivity of 57.14% (48/84) and a specificity of 93.4% (127/136) compared with 60.7% and 94.9% when using the conventional program on the same set of data.

A comparison of the discrete and smoothed methods will largely depend on the difference of the sensitivities and the specificities and on the number of ECGs in the population. However, further consideration must be given to the level of agreement between each method on the numbers of true positives (TP), false negatives (FN), true negatives (TN) and false positives (FP) thus assessing the correlation of the two methods (Bailey et. al., 1988).

For any comparison of two distinct methods on each of N individuals a table showing the agreement of classification can be constructed as follows:-

		METHOD 2		
		TP	FN	TOTAL
METHOD 1	TP	A	B	N_1
	FN	C	D	$N-N_1$
	TOTAL	N_2	$N-N_2$	N

Table 6.2 **Comparison of sensitivity between two methods.**

where A = number of TP for both method 1 and method 2

B = number of TP for method 1 and FN for method 2

C = number of FN for method 1 and TP for method 2

D = number of FN for method 1 and method 2.

The sensitivities (and specificities) of method 1 and method 2 (S_1 and S_2) may be compared using McNemar's test based on the

'disagreements' between the two methods, i.e. based on the values of B and C. McNemar (1947) states that, under the null hypothesis that the sensitivities of each method are equal (i.e. $S_1 = S_2$), the test statistic X^2 is approximated by a χ^2 distribution with one degree of freedom where

$$X^2 = \frac{(B-C)^2}{B+C}$$

and B and C are as defined above.

At the 5% level it can therefore be deduced that S_1 and S_2 will differ significantly if $X^2 > 3.84$.

Table 6.3 demonstrates the level of agreement in terms of true positives and false negatives between the conventional 'discrete' program (method 1) and the modified 'smoothed' program (method 2).

		Smoothed Method		
		TP	FN	TOTAL
Discrete	TP	46	5	51
Method	FN	2	31	33
	TOTAL	48	36	84

Table 6.3 **Comparison of the sensitivities of the Discrete and Smoothed programs.**

In this case $X^2 = 1.28$. Therefore there is no evidence to suggest that there is a significant difference in sensitivity between the conventional method and the new smooth approach.

Similarly, agreement between true negatives and false positives can be tabulated and a comparison of the specificity of each method made.

		Smoothed Method		
		TN	FP	TOTAL
Discrete Method	TN	125	4	129
	FP	2	5	7
	TOTAL	127	9	136

Table 6.4 **Comparison of the specificities of the Discrete and Smoothed programs**

In this case, $X^2 = 0.667$. Again, there is no evidence to suggest that there is a significant difference in specificity between the two methods.

6.6 COMPARING REPEATABILITY OF THE EXISTING PROGRAM AND THE MODIFIED PROGRAM

6.6.1 Day-to-day ECG Recordings

Since there is no difference in the diagnostic performance of the existing program and the smoothed version we can now proceed to the key question of interest and compare the repeatability of both methods. Diagnostic interpretations for consecutive ECG recordings using the smoothed version of the program are expected to be in closer agreement than those obtained using the existing method. This is largely due to the fact that diagnostic indices will be assigned continuously instead of on a discrete basis. This means that small measurement changes (even in the vicinity of threshold values) should result only in small differences in the scores produced. Thus, comparable measurements which previously crossed threshold values from one day to the next resulting in different scores using the

existing program should now receive similar diagnostic indices depending on their proximity to the critical value and on the estimated amount of day-to-day variability in that particular measurement.

The indices associated with Left Ventricular Hypertrophy (LVH scores) were obtained using both the existing and the modified versions of the program on a group of 330 patients who were admitted to Glasgow Royal Infirmary between August 1988 and December 1991 and were not suffering from any acute cardiac illness. The exact diagnosis does not matter in stable patients given that it is the repeatability of the method that is being tested. Subsequent ECGs were recorded at least 24 hours later to enable the repeatability of the techniques to be assessed. The Glasgow program classifies individuals as having no LVH if the resulting diagnostic index (LVH Score) is 3 or less and as having LVH of one form or another if it is greater than or equal to 4.

i.e.

If LVH Score $\leq$ 3 then ' No LVH '

$\geq$ 4 then ' LVH '

Using the 'discrete' version of the Glasgow program applied separately to each patient on each of the 2 days produced the results in Table 6.5.

DISCRETE CASE

		DAY1	
		NOLVH	LVH
DAY2	NOLVH	273	17
	LVH		40

Table 6.5 **Repeatability of existing method. (Day-to-day ECG recordings)**

Seventeen individuals therefore have inconsistent LVH diagnoses from one day to the next when using the discrete version of the Glasgow program. Note that the number in the top right hand cell in Table 6.5 corresponds to the sum of the off-diagonal elements since day 1 and day 2 are essentially interchangeable. In the tables that follow, the same approach has been adopted.

Applying the new smoothed program with a cutpoint of 3.5 previously suggested to be 'optimal' produces Table 6.6 which reduces the number of inconsistent diagnoses to 9 - an almost 50% decrease on the discrete version.

SMOOTHED CASE

		DAY 1	
		NO LVH	LVH
DAY 2	NO LVH	277	9
	LVH		44

Table 6.6 **Repeatability of smoothed method. (Day-to-day ECG recordings)**

Since the smoothed version produces a continuous score rather than a discrete score it is of interest to compare the close agreement between the smooth day 1 LVH scores and the smooth day 2 LVH scores as illustrated in Figure 6.8. All of the points appear to lie along a line with approximately unit slope. It can be seen that of the nine day-to-day recordings which were inconsistent in the diagnosis of LVH, most of the LVH scores were located near to the boundary. However, there is one point which exhibits a substantial difference in LVH score from day to day. On day 1 the smooth LVH score is 6.1 and on day 2 the score is 3.2. The corresponding discrete scores are 6 and 3.

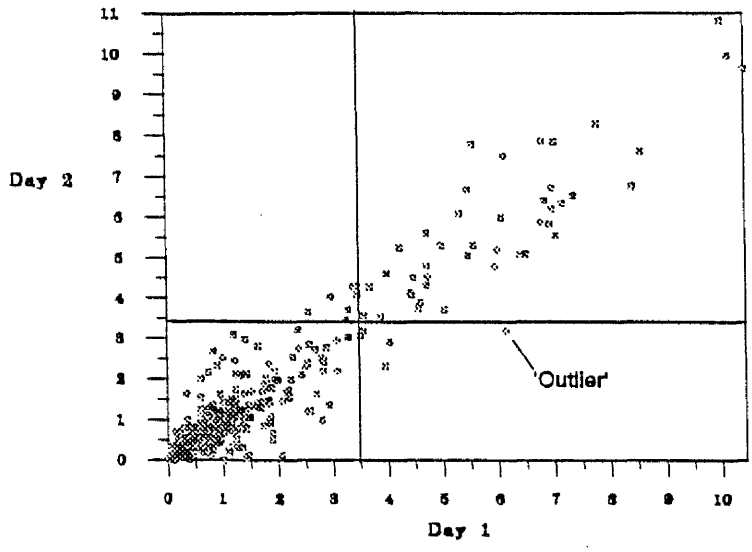


Figure 6.8 **Plot of Day1 scores vs. Day2 scores**

On closer inspection it appears that the individual in question has an abnormally high Lewis Index on day 1 but not on day 2.

As an extension to merely detecting the presence or absence of LVH, the Glasgow program allows the degree of LVH to be assessed. The degree to which a patient exhibits LVH is assessed on the basis of the magnitude of the LVH Score in the following way:

If	LVH Score	≤	3	then	' No LVH '
		=	4	then	' Possible LVH '
		=	5	then	' Probable LVH '
		≥	6	then	' Definite LVH '

It is of substantive interest to assess the severity of the inconsistent interpretations given that a day 1 diagnosis of 'No LVH' followed by a day 2 diagnosis of 'Definite LVH' is less acceptable than 'No LVH' followed by 'Possible LVH'. In order to do this, cutpoints separating Possible from Probable LVH and Probable from Definite LVH for the modified program need to be selected.

Table 6.7 illustrates the repeatability of the existing program when the diagnosis of LVH is grouped into the four possible categories (No LVH, Possible, Probable and Definite LVH).

DISCRETE CASE

		DAY1			
		NO LVH	POSS	PROB	DEF
DAY 2	NO LVH	273	8	5	4
	POSS		4	5	2
	PROB			2	9
	DEF				18

Table 6.7 **Repeatability of existing method when comparing the severity of LVH (Day-to-day ECG recordings)**

A simple way of assessing how 'repeatable' the program is can be achieved by assigning a 'Repeatability Index' as follows :

Score 0 for each pair of ECGs where the diagnosis is the same on both occasions

Score 1 if the diagnosis changes by one category only (i.e. No LVH → Poss., Prob → Def. etc.)

Score 2 if the diagnosis changes by two categories (i.e. No LVH → Prob., Poss → Def.)

Score 3 if the diagnosis changes by three categories (i.e. No LVH → Def.)

One way of evaluating the performance of the smoothing technique is to calculate this 'Repeatability Index' (RI) over all individuals and use the sum of such values as an overall criterion of performance.

Using the above scoring technique, the existing program has an overall 'Repeatability Index' of 48

(i.e. $0\times(273+4+2+18)+1\times(8+5+9)+2\times(5+2)+3\times(4)$)

based on the 330 patients.

In order that a comparison could be made between the existing and modified programs, cutpoints separating the different categories of hypertrophy for the smooth scoring system were required. Initially it was thought that the relationship between smooth LVH score and left ventricular mass for the 84 clinically documented LVH cases could be explored as a method of selecting optimal cutpoints. However, problems arose due to the fact that left ventricular mass and subsequent identification of LVH is sex-dependent. Instead, a method based on the smooth LVH scores from the day 1 ECG recordings was devised. In Figure 6.9 the four various discrete classifications of LVH for the day 1 ECG recordings have been used to produce four similar frequency distributions for the smooth score.

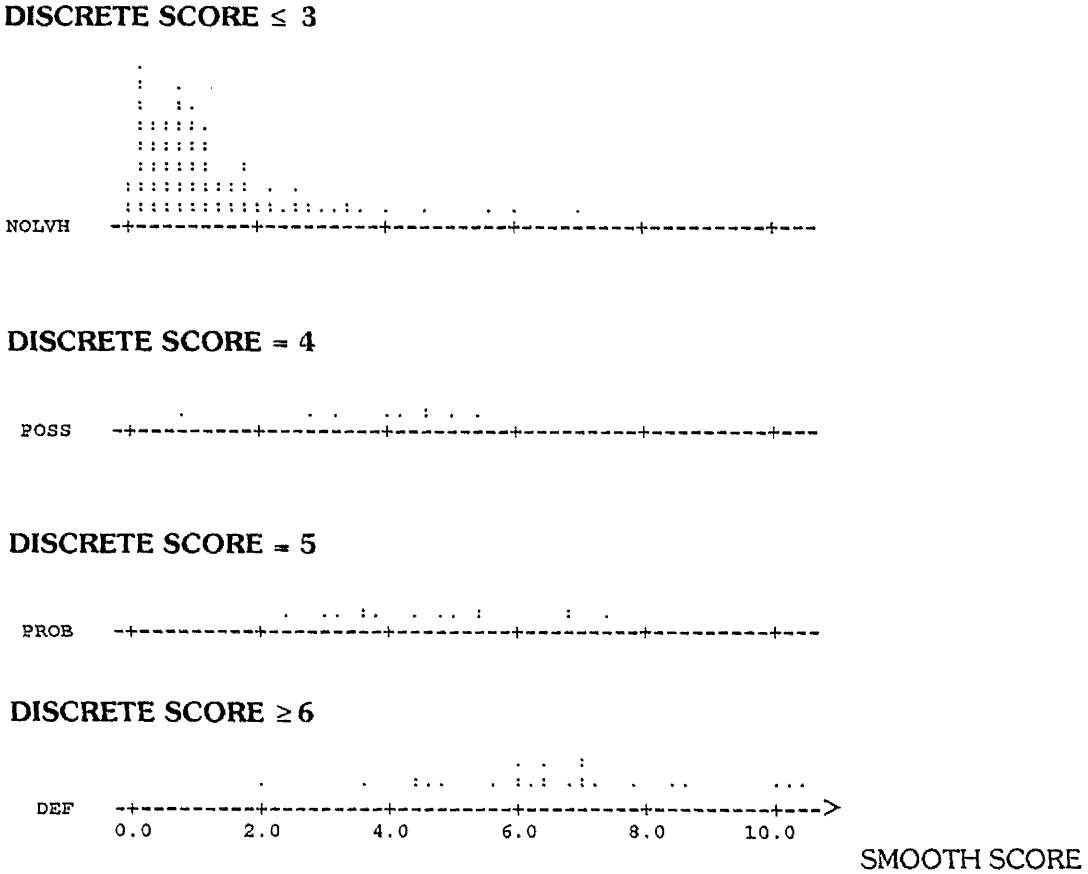


Figure 6.9 **Frequency Distribution of the smoothed LV score (Day 1 recordings)**

Values of 3.5, 4.7 and 6.0 were selected as being the cutpoints which best separate out the four categories of LVH as described by the corresponding day 1 discrete values. These values produce the following table which results in a 'Repeatability Index' of 23.

SMOOTHED CASE

		DAY1			
		NO LVH	POSS	PROB	DEF
DAY 2	NO LVH	277	8	0	1
	POSS		10	4	0
	PROB			6	8
	DEF				16

Table 6.8 **Repeatability of smoothed method when comparing the severity of LVH (Day-to-day ECG recordings)**

Of the 22 diagnoses which were inconsistent by one category using the existing program, 16 are now in agreement using the modified program and 6 remain the same (see Table 6.9). Similarly, 5 of the 7 slightly more serious inconsistent classifications are now in agreement while the remaining 2 are inconsistent by one category only. Of the 4 most serious discrepancies (i.e. normal on one occasion with definite LVH on the other), 2 now are in agreement, the degree of inconsistency has decreased in one and the last one remains the same.

		No. of Discrepancies (Discrete Method)			
		0	1	2	3
No. of Discrepancies (Smooth Method)	0	286	16	5	2
	1	11	6	2	1
	2	0	0	0	0
	3	0	0	0	1

Table 6.9 **Comparing the inconsistent diagnoses of the existing and the smoothed methods. (Day-to-day ECG recordings)**

Interestingly, 11 pairs of ECG recordings which displayed no differences in terms of the diagnosis of LVH using the discrete method now appear as inconsistent (i.e. normal → possible LVH or vice versa) using the modified version of the program. However, in all eleven replicates there is one recording which produces an LVH score which is close to one of the arbitrary boundaries described previously. Slight alteration of the boundary values 3.5, 4.7 and 6.0 may improve the repeatability of these 11 pairs of ECGs. However, this would probably produce a further set of inconsistent ECGs with LVH scores close to the new boundary values.

In general, the smooth method appears to be far superior to the conventional method in terms of repeatability of the day-to-day diagnosis of LVH. It has been demonstrated that 23 of the cases which displayed some form of change with the discrete method were in agreement when the smooth techniques were applied. Of these 23 cases, 2 originally displayed serious discrepancies (normal → definite LVH or vice versa) and 5 showed moderate changes. In contrast, only 11 of the 297 pairs of ECGs which were totally in agreement when the conventional method was used displayed changes when the

smooth method was used, all of these changes being of a relatively minor importance (i.e. normal → possible LVH or vice versa).

6.6.2 Minute-to-minute ECG Recordings.

It is also of interest to establish whether there was a significant lack of repeatability in terms of LVH diagnosis between 249 pairs of ECGs which had been recorded one minute apart. Any inconsistencies arising could not be caused by misplaced electrodes or by between-technician variation and would therefore be attributed to other sources of variability, mainly computer processing techniques.

Using the conventional method, only 8 pairs of ECGs produced incompatible LVH/No LVH diagnoses (Table 6.10) compared with one set of ECGs when the modified version was used (Table 6.11). This is clearly a significant improvement, showing that the new approach virtually abolishes repeat changes with respect to LVH/No LVH.

DISCRETE CASE

		MIN 1	
		NO LVH	LVH
MIN 2	NO LVH	211	8
	LVH		30

Table 6.10 **Repeatability of the existing method (Minute-to-minute ECG recordings)**

SMOOTHED CASE

		MIN 1	
		NO LVH	LVH
MIN 2	NO LVH	210	1
	LVH		38

Table 6.11 **Repeatability of the smooth method (Minute-to-minute ECG recordings)**

Figure 6.10 demonstrates the close agreement between the minute 1 LVH scores and the minute 2 LVH scores. The plot also illustrates that the scores corresponding to the single pair of minute-to-minute ECG recordings which produced an inconsistent diagnosis are situated close to the boundary value of 3.5.

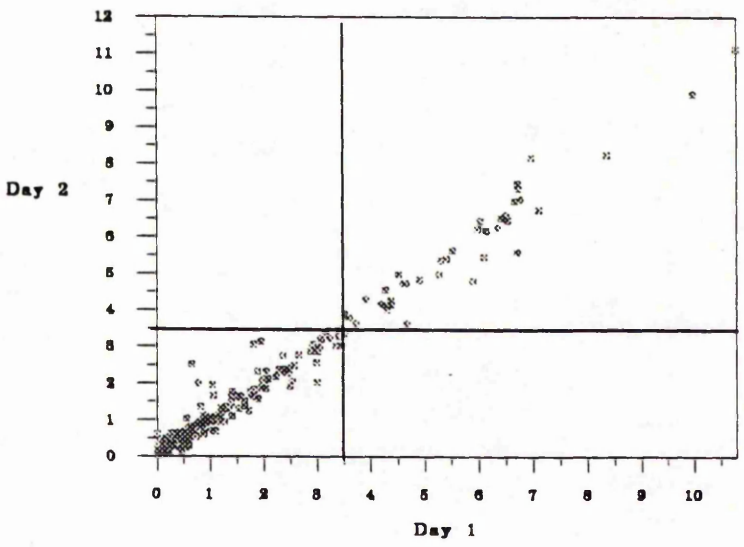
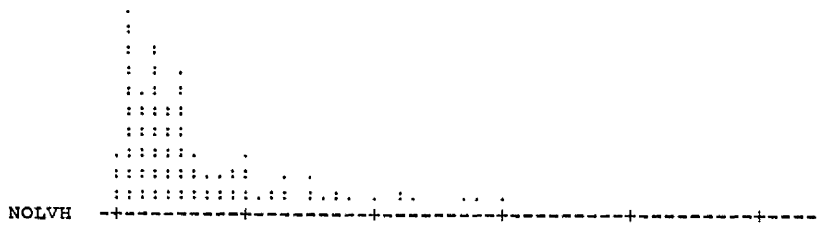


Figure 6.10 **Plot of Minute 1 scores vs. Minute 2 scores**

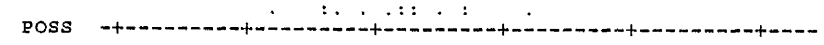
Figure 6.11 illustrates the separation of the four different classifications of LVH on the basis of the existing diagnosis for the day 1 recording.

Table 6.12 indicates that, when using the conventional version of the Glasgow program, there are 13 discrepancies compared to only 6 when the modified version (Table 6.13) is used.

DISCRETE SCORE ≤ 3



DISCRETE SCORE =4



DISCRETE SCORE =5

DISCRETE SCORE ≥ 6

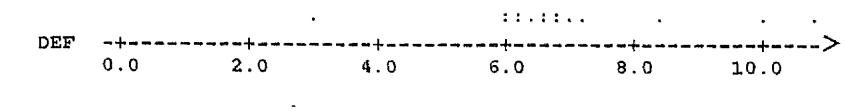


Figure 6.11 **Frequency Distribution of smoothed LV score (Minute 1 ECGs)**

DISCRETE CASE

		MIN1			
		NO LVH	POSS	PROB	DEF
MIN 2	NO LVH	211	6	2	0
	POSS		9	1	1
	PROB			2	3
	DEF				14

Table 6.12 **Repeatability of the existing method when comparing the severity of LVH (Minute-to-minute ECG recordings)**

SMOOTHED CASE

		MIN1			
		NO LVH	POSS	PROB	DEF
MIN 2	NO LVH	210	1	0	0
	POSS		11	3	0
	PROB			6	2
	DEF				16

Table 6.13 **Repeatability of smoothed method when comparing the severity of LVH (Minute-to-minute ECG recordings)**

The Repeatability index drops from 16 to 6 when the modified version of the program is used.

Table 6.14 illustrates the relationship between the discrepancies for each method. There are 4 pairs of ECG recordings which produced the same diagnoses using the discrete method but demonstrate inconsistencies when the smoothed version is used. Again this is attributable to the fact that the LVH scores lie close to the previously suggested cutpoints.

		No. of Discrepancies (Discrete Method)			
		0	1	2	3
No. of Discrepancies (Smooth Method)	0	232	8	3	0
	1	4	2	0	0
	2	0	0	0	0
	3	0	0	0	0

Table 6.14 **Comparing the inconsistent diagnoses of the existing and the smoothed methods. (Minute-to-minute ECG recordings)**

6.6.3 Split ECG Recordings.

A method for testing the repeatability of computer programs for ECG interpretation was established by Bailey and his colleagues (Bailey et. al., 1974). Instead of examining two ECGs for each individual, two digital representations of the same tracing were obtained, each being separated by one millisecond in time. In this classic publication, initial ECGs were recorded with a sampling rate of 1000 samples per second. Extraction of the odd and the even samples then provided two separate tracings, each representing the digital data at 500 samples per second.

In a similar way, two tracings were obtained from each of the 330 day 1 recordings which have been described in section 6.9. These tracings represent the digital data at 250 samples per second since the initial ECGs were digitised at a sampling rate of 500. The tracings were then interpolated back to a sampling rate of 500 for the purpose of analysis. The repeatability of the diagnosis of LVH was then assessed for both the existing method and the smooth method.

Tables 6.15 and 6.16 demonstrate that the agreement between two representations of the same ECG is superior to the agreement between ECGs recorded either one day apart or within one minute of each other. When the existing method is used there are 4 inconsistent diagnoses compared to 2 when the smooth method is used.

DISCRETE CASE

		ECG 1	
		NO LVH	LVH
ECG 2	NO LVH	283	4
	LVH		43

Table 6.15 **Repeatability of the existing method (Split ECG recordings)**

SMOOTHED CASE

		ECG 1	
		NO LVH	LVH
ECG 2	NO LVH	283	2
	LVH		45

Table 6.16 **Repeatability of the smooth method (Split ECG recordings)**

However, when considering the subtleties of the diagnostic statements in terms of Possible, Probable and Definite LVH there is still a slight reduction in the Repeatability Index from 9 using the existing method (Table 6.17) to 5 when the smooth method is used (Table 6.18).

DISCRETE CASE

		ECG1			
		NO LVH	POSS	PROB	DEF
ECG 2	NO LVH	283	2	2	0
	POSS		9	0	0
	PROB			8	3
	DEF				23

Table 6.17 **Repeatability of the existing method when comparing the severity of LVH (Split ECG recordings)**

SMOOTHED CASE

		ECG1			
		NO LVH	POSS	PROB	DEF
ECG 2	NO LVH	283	2	0	0
	POSS		13	1	0
	PROB			8	2
	DEF				21

Table 6.18 **Repeatability of the smooth method when comparing the severity of LVH (Split ECG recordings).**

6.7 SUMMARY

In an attempt to improve the repeatability of the section of the diagnostic program associated with Left Ventricular Hypertrophy in use in the Glasgow laboratory, discrete upper values of 'healthiness' were replaced with continuously changing limits where appropriate. This to a considerable extent alleviated the problem of inconsistent interpretations occurring as a result of, for example, a patient's age

group changing from one recording to the next or a slight change in a particular ECG measurement. Furthermore, the discrete index initially associated with the final diagnosis of LVH was replaced by a smoothed version. This allowed measurements which lay close to (but below) threshold values to receive a proportion of the relevant contribution to the final score, this proportion depending on the amount of day-to-day variation inherent in that particular measurement.

The new smoothed diagnostic indices were calculated for a group of 84 clinically documented LVH cases and 136 non-LVH cases in order to compare the sensitivity and specificity of the new smoothed version with the existing method. No significant differences were found indicating that the diagnostic accuracy for each method was comparable.

However, examination of the repeatability of each method revealed that the smoothed version was far superior. Based on a sample of 330 pairs of ECGs which were recorded at least 24 hours apart, 33 pairs produced inconsistent diagnoses of LVH using the existing method. This number dropped to 21 when the smoothed version was used. The greatest improvement in repeatability was in terms of the severity of the inconsistencies. Using the existing method 22 discrepancies were of a mild nature, 7 were moderate and 4 were severe whereas the smoothed version produced 20, 0 and 1 respectively.

Similar results were seen when examining the repeatability of the minute-to-minute ECG recordings. Of 249 pairs of ECGs which were recorded only one minute apart (without removing and subsequently replacing the electrodes), 13 pairs produced inconsistent diagnoses using the existing method compared to only 6 with the smoothed

version. Again the inconsistencies using the existing method were more severe - 10 mild and 3 moderate compared to only 6 mild using the smoothed method.

Further examination of the repeatability of both the existing and the smooth methods for diagnosing LVH was possible by making use of a technique established by Bailey et. al. (1974). This approach involved splitting one digital representation of an ECG tracing recorded at 500 samples per second into two separate tracings, one representing the 250 odd samples, the other representing the 250 even samples. Of the 330 day 1 ECG recordings which were split in this way there were 7 inconsistent diagnoses using the existing method, 5 of which were mild, and 2 moderate, compared to only 5 mild discrepancies using the smooth method.

There will be situations when differences in ECG measurements from one recording to the next will be so large that no amount of smoothing will solve the problem of lack of repeatability. However, it is clear that the application of smoothing techniques vastly improves the repeatability of the diagnosis of LVH in the Glasgow program whilst preserving its diagnostic accuracy.

CHAPTER SEVEN:

ST-T CHANGES.

7.1 INTRODUCTION

Some of the difficulties encountered when attempting to diagnose abnormalities of the ST-T segment were outlined in Chapter 5 while Chapter 6 examined some of the complexities involved when diagnosing left ventricular strain in the presence of left ventricular hypertrophy.

Routinely, diagnoses relating to abnormalities of the ST segment are made on the basis of both the magnitude of the ST amplitude and the slope of the ST segment. There are potentially many opportunities for discrete thresholds to be crossed from one recording to the next which in turn may result in changes in diagnoses from one recording to the next. By applying the smoothing techniques which have already been described, we hope to improve the repeatability of the section of the Glasgow program which deals with ST abnormalities.

The techniques which have been described in previous chapters can be used in one-dimensional and multi-dimensional situations and have been extended to cater for the situation where several steps are used to assess the severity of a certain condition.

Since diagnosis of ST changes in the inferior leads is seldom reported in isolation and is usually found in combination with other conditions such as T wave abnormalities accompanying LVH and myocardial infarction, we will direct attention to the level of

agreement in the actual recording-to-recording indices of ST depression in the inferior leads instead of the diagnostic statements.

7.2 COMPARISON OF DISCRETE AND SMOOTHED DIAGNOSTIC INDICES (ST Depression).

7.2.1 Day-to-day ECG Recordings.

In the Glasgow program, depression of the ST segment is described as being 'Equivocal', 'Moderate', or 'Marked' depending on the magnitude of a discrete ST score $S(x_1, x_2)$. This score is calculated on the basis of two measurements, namely the negative of the ST amplitude x_1 and the ST slope x_2, in the following way:

if	$x_1 < -100\mu V$	and $x_2 < 0°$	then	$S(x_1, x_2) = -3$
	$x_1 < -50\mu V$	and $x_2 < 0°$	then	$S(x_1, x_2) = -2$
	$x_1 < -20\mu V$	and $x_2 < 0°$	then	$S(x_1, x_2) = -1$
	$x_1 > 60\mu V$	and $x_2 > 0°$	then	$S(x_1, x_2) = 1$
	$x_1 > 80\mu V$	and $x_2 > 0°$	then	$S(x_1, x_2) = 2$
	$x_1 > 100\mu V$		then	$S(x_1, x_2) = 3$
	otherwise			$S(x_1, x_2) = 0$

The statements are then assigned as follows:

If	$S(x_1, x_2)$	= 0	then	'No ST Depression'
		= -1	then	'Equivocal ST Depression'
		= -2	then	'Moderate ST Depression'
		= -3	then	'Marked ST Depression'

Initially, although the 'true' amount of ST depression remains unknown, a comparison was made between the smooth scores and the discrete scores which were obtained for the index of ST depression in lead II from the initial recordings of our sample of 330 day 1 and day 2 ECGs. These new smooth scores were based on the smoothed ST scores in the appropriate leads and a small section of the relevant Fortran code can be seen in part B) of the Appendix. Although we are primarily interested in the amount of day-to-day variability in the diagnoses of ST abnormalities, we first wish to establish that there are no fundamental differences between the smooth and conventional methods of assigning ST indices. Since the smooth method of diagnosing abnormalities of the ST segment has been devised in an attempt to improve repeatability from recording to recording, we do not expect this smoothed approach to differ dramatically from the existing method when applied to a set of single ECG recordings.

The smoothed version of the diagnostic index for ST depression is now continuous in nature (see Chapter 5) so that acceptable values for cutpoints which separate the various categories of ST depression are now required. A dot plot of the smoothed scores for each of the 4 values of the discrete scores for a training set of 151 of the day 1 ECGs is presented in Figure 7.1.

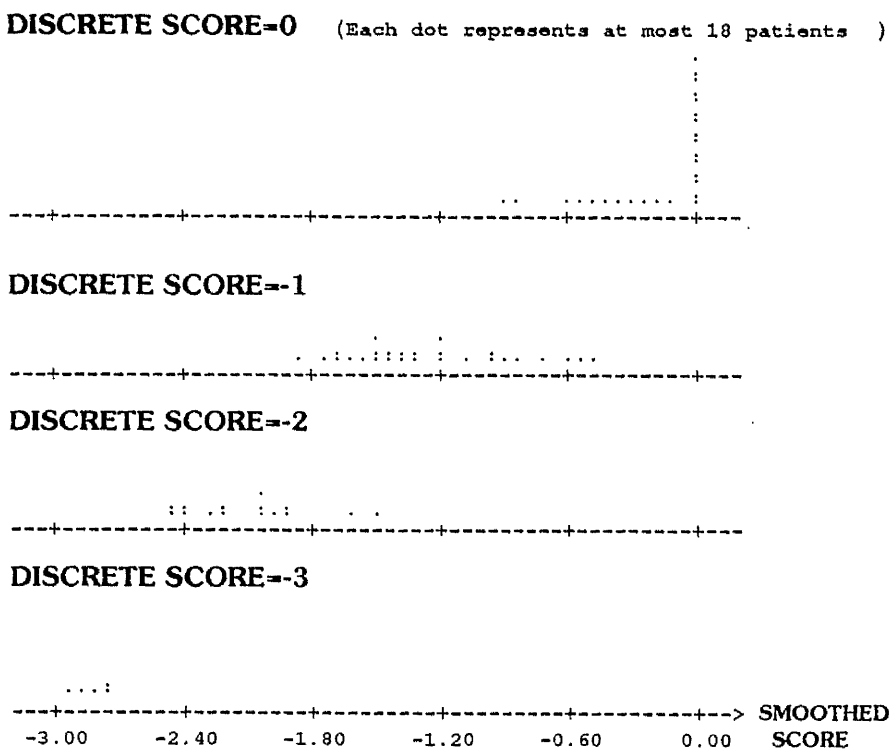


Figure 7.1 **Frequency Distribution of the smoothed ST Depression Index (Day 1 recordings)**

In order to compare the repeatability of the smooth approach of diagnosing abnormalities of the ST segment with the conventional method we must provide a continuous analogue of the discrete categorisation. Cutpoint values of -0.9, -1.8 and -2.7 were chosen in order to separate the four categories of ST depression on the basis of the smoothed scores. The categorisation was achieved in the following way:

If	$-0.9 \leq S_m^*(x_1, x_2) < 0$	then	'No ST Depression'
	$-1.8 \leq S_m^*(x_1, x_2) < -0.9$	then	'Equivocal ST Depression'
	$-2.7 \leq S_m^*(x_1, x_2) < -1.8$	then	'Moderate ST Depression'
	$-3.0 \leq S_m^*(x_1, x_2) < -2.7$	then	'Marked ST Depression'

Table 7.1 illustrates the level of agreement between the smooth and the discrete approaches on the basis of these cutpoints.

		DISCRETE	STDEP	DAY1	(II)
		NONE	EQUIV	MOD	MARK
SMOOTH	NONE	286	2	0	0
STDEP	EQUIV	4	20	1	0
DAY1	MOD	0	3	10	0
(II)	MARK	0	0	0	4

Table 7.1 **A comparison of the existing and the smooth methods of diagnosing ST Depression in lead II (Day-to-day ECG recordings).**

In order to make a comparison of the repeatability of the two methods of detecting ST changes, ST scores were obtained for consecutive recordings taken at least 24 hours apart from the same individuals. For presentation purposes, only the ST scores in the inferior leads (II and aVF) are tabulated although it should be noted that comparisons may be made for all twelve leads.

Table 7.2 illustrates the level of agreement of the ST depression indices between day 1 and day 2 recordings for the existing program for lead II.

DISCRETE CASE

		STDEP	DAY1	(II)	
		NONE	EQUIV	MOD	MARK
STDEP	NONE	274	19	3	0
DAY2	EQUIV		13	8	0
(II)	MOD			8	2
	MARK				3

Table 7.2 **Repeatability of the existing method when comparing the severity of ST Depression in lead II (Day-to-day ECG recordings).**

There are 274+13+8+3=298 of the 330 repeat ECGs exhibiting no change from day 1 to day 2, 19+8+2=29 cases having a 'mild' discrepancy and 3 cases with 'moderate' discrepancies from recording to recording. The corresponding 'Repeatability Index' (RI) is 35.

In contrast, the Table 7.3 shows the results obtained when using the modified program.

SMOOTHED CASE

		STDEP DAY1 (II)			
		NONE	EQUIV	MOD	MARK
STDEP	NONE	286	11	1	0
DAY2	EQUIV		14	4	0
(II)	MOD			9	2
	MARK				3

Table 7.3 **Repeatability of the smooth method when comparing the severity of ST Depression in lead II (Day-to-day ECG recordings).**

By comparison, there are 286+14+9+3=312 cases with no change from day 1 to day 2, 11+4+2=17 cases with 'mild' discrepancies and 1 case with a 'moderate' discrepancy. The 'Repeatability Index' has been reduced from 35 to 19, demonstrating a considerable improvement in terms of repeatability.

It is interesting to discover whether the 17 cases showing a mild discrepancy with the modified program are themselves a subset of the 29 obtained when using the existing program. To achieve this, Table 7.4 has been constructed. From this Table it is easy to see that of these 29 cases, 14 remain in the same category while the remaining 15 cases show no form of discrepancy in ST Depression Index from

one day to the next. Thus, for this category, inconsistencies have been halved in number using the new approach.

No. of Discrepancies (Smooth Method)	No. of Discrepancies (Discrete Method) 0	1	2
0	297	15	0
1	1	14	2
2	0	0	1

Table 7.4 **Comparing the inconsistent diagnoses of the existing and smoothed methods for the day-to-day ECG recordings (lead II).**

Table 7.4 also draws attention to the fact that one of the cases is worse in terms of repeatability when the modified program is used. This is an unexpected occurrence. On closer examination however, it turns out that the smooth day 1 score of ST depression in lead II is -1.49 and the day 2 score is -0.82. This second value is very close to the cutpoint of -0.9 which has been suggested for separating those with no ST depression from those with equivocal ST depression.

Similar tables were constructed for the ST depression indices in lead aVF. Comparisons between the day 1 and day 2 indices can be seen in Tables 7.5 and 7.6.

DISCRETE CASE

		STDEP DAY1 (aVF) NONE	EQUIV	MOD	MARK
	NONE	276	21	1	0
DAY2	EQUIV		20	4	0
(aVF)	MOD			6	1
	MARK				1

Table 7.5 **Repeatability of the existing method when comparing the severity of ST Depression in lead aVF (Day-to-day ECG recordings).**

SMOOTHED CASE

		STDEP NONE	DAY1 EQUIV	(aVF) MOD	MARK
STDEP	NONE	288	14	0	0
DAY2	EQUIV		15	4	0
(aVF)	MOD			8	0
	MARK				1

Table 7.6 **Repeatability of the smooth method when comparing the severity of ST Depression in lead aVF (Day-to-day ECG recordings).**

The number of 'mild' discrepancies has dropped from 26 to 18 and the 'moderate' discrepancy has been eliminated.

		No. of Discrepancies (Discrete Method)		
		0	1	2
No. of Discrepancies (Smooth Method)	0	300	12	0
	1	3	14	1
	2	0	0	0

Table 7.7 **Comparing the inconsistent diagnoses of the existing and smoothed methods for day-to-day ECG recordings (lead aVF).**

From Table 7.7 it can be seen that three of the cases are worse in terms of repeatability when the modified program is used. However, it appears that at least one of the recordings from each pair of ECGs produces an ST depression index which is close to a threshold value.

The repeatability indices for Tables 7.5 and 7.6 are 28 and 18 respectively, again showing an improvement in terms of day to day

agreement between ST scores for the inferior lead aVF when the new approach is used.

7.2.2 Minute-to-minute ECG Recordings.

It was also of interest to discover whether there was a significant improvement in the minute-to-minute repeatability of the calculation of Inferior ST Depression indices. To investigate this, two recordings were taken from individuals without removing and subsequently replacing the electrodes. It was expected that inconsistencies in interpretations obtained from these minute-to-minute recordings would be of a less severe nature than those observed from day to day since variation due to electrode positioning is being eliminated.

Table 7.8 demonstrates that 9 pairs of 249 ECGs which were recorded a minute apart produce inconsistent ST indices in lead II, resulting in a 'Repeatability Index' of 12. The modified version of the program eliminates one of these inconsistencies, the 'Repeatability Index' dropping to 11 (see Table 7.9).

DISCRETE CASE

		STDEP	DAY1	(II)	
		NONE	EQUIV	MOD	MARK
STDEP	NONE	224	5	1	1
DAY2	EQUIV		10	2	0
(II)	MOD			4	0
	MARK				2

Table 7.8 **Repeatability of the existing method when comparing the severity of ST Depression in lead II (Minute-to-minute ECG recordings).**

SMOOTHED CASE

		STDEP NONE	DAY1 EQUIV	(II) MOD	MARK
STDEP	NONE	230	5	1	1
DAY2	EQUIV		6	0	0
(II)	MOD			4	1
	MARK				0

Table 7.9 **Repeatability of the smooth method when comparing the severity of ST Depression in lead II (Minute-to-minute ECG recordings).**

Tables 7.10 and 7.11 demonstrate that the 'Repeatability Index' is reduced from 13 to 6 when the smoothed techniques are used to calculate the ST Depression indices from minute to minute recordings in lead aVF which certainly represents a substantial improvement.

DISCRETE CASE

		STDEP NONE	MIN1 EQUIV	(aVF) MOD	MARK
STDEP	NONE	225	10	0	0
MIN2	EQUIV		8	3	0
(aVF)	MOD			3	0
	MARK				0

Table 7.10 **Repeatability of the existing method when comparing the severity of ST Depression in lead aVF (Minute-to-minute ECG recordings).**

SMOOTHED CASE

		STDEP NONE	MIN1 EQUIV	(aVF) MOD	MARK
STDEP	NONE	231	5	0	0
MIN2	EQUIV		9	1	0
(aVF)	MOD			3	0
	MARK				0

Table 7.11 Repeatability of the smooth method when comparing the severity of ST Depression in lead aVF (Minute-to-minute ECG recordings).

7.2.3 Split ECG Recordings

A third method of assessing the repeatability of the diagnosis of ST-T changes of both the existing and the smooth methods is possible if we split the day 1 ECGs which have been sampled at a rate of 500 samples per second into two tracings which represent the data sampled at a rate of 250. ST indices derived using both the discrete and the smooth methods were obtained from the 330 day 1 ECG tracings in this way. Table 7.12 demonstrates that the 'Repeatability Index' for the existing method applied to lead II is 9. When the smooth method is used, the RI increases slightly to 11 (see Table 7.13). This observed increase in 'Repeatability Index' is inconsistent with our previous findings. Although there are a further two pairs of ECGs which exhibit discrepant ST indices in lead II when the smooth method is used the indices which are calculated are very similar. Unfortunately, in the first case, the indices are -0.89 and -0.91 which lie on either side of the cutpoint of -0.9 which is used to separate 'No ST depression' from 'Equivocal ST depression'. Similarly, the second pair of ECGs produces indices of -1.81 and

-1.79. Again these are very close to a cutpoint, this time the value of -1.8 which is used to separate 'Equivocal ST depression' from 'Moderate ST depression'.

DISCRETE CASE

		STDEP NONE	MIN1 EQUIV	(II) MOD	MARK
STDEP	NONE	284	6	0	0
MIN2	EQUIV		22	3	0
(II)	MOD			12	0
	MARK				3

Table 7.12 **Repeatability of the existing method when comparing the severity of ST Depression in lead II (Split ECG recordings).**

SMOOTHED CASE

		STDEP NONE	MIN1 EQUIV	(II) MOD	MARK
STDEP	NONE	287	6	0	0
MIN2	EQUIV		18	5	0
(II)	MOD			11	0
	MARK				3

Table 7.13 **Repeatability of the smooth method when comparing the severity of ST Depression in lead II (Split ECG recordings).**

However, the improvement due to smoothing returns in lead aVF. In this case, the 'Repeatability Index' drops from 14 to 9 when the discrete method is replaced by the smooth method (see Tables 7.14 and 7.15).

DISCRETE CASE

		STDEP NONE	MIN1 EQUIV	(aVF) MOD	MARK
STDEP	NONE	285	12	1	0
MIN2	EQUIV		21	0	0
(aVF)	MOD			10	0
	MARK				1

Table 7.14 **Repeatability of the existing method when comparing the severity of ST Depression in lead aVF (Split ECG recordings).**

SMOOTHED CASE

		STDEP NONE	MIN1 EQUIV	(aVF) MOD	MARK
STDEP	NONE	294	7	0	0
MIN2	EQUIV		14	2	0
(aVF)	MOD			12	0
	MARK				1

Table 7.15 **Repeatability of the smooth method when comparing the severity of ST Depression in lead aVF (Split ECG recordings).**

7.3 RE-DEFINING ST DEPRESSION AND ST ELEVATION IN THE INFERIOR LEADS

Once the ST indices described in Chapter 5 have been calculated for each lead individually, these scores are examined for evidence of ST elevation or ST depression. Instead of providing a measure of depression or elevation for every lead, leads II and aVF are grouped to supply an index relating to the inferior leads and similarly for the lateral, anterior, anteroseptal and septal leads.

7.3.1 ST Depression

We may now combine the smooth ST indices which have been calculated for leads II and aVF separately in an attempt to enhance the repeatability of the diagnosis of inferior ST depression. Currently, ST depression in the inferior leads is diagnosed if the minimum of the indices in leads II and aVF is a certain value. The lower this value the more marked the depression,

i.e. In either lead

if $S(x_1, x_2)$ = -1 then STDPIN = -1

if $S(x_1, x_2)$ = -2 then STDPIN = -2

if $S(x_1, x_2)$ = -3 then STDPIN = -3

where STDPIN denotes the index of ST depression in the Inferior leads II and aVF. This scheme is represented diagramatically in Figure 7.2.

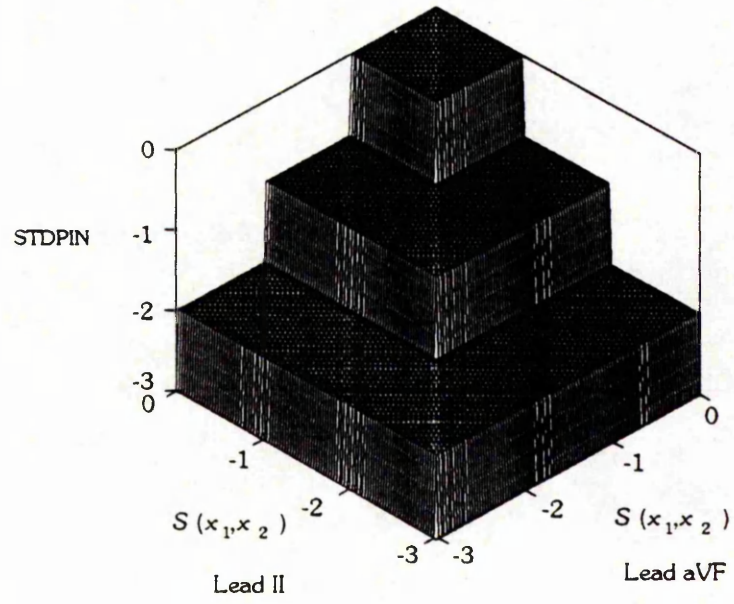


Figure 7.2 **Diagrammatic representation of discrete score function for Inferior ST Depression.**

Table 7.16 compares the level of agreement of the Inferior ST Depression statements for pairs of ECGs which have been recorded on consecutive days. There are 40 mild and 3 moderate inconsistencies, producing a 'Repeatability Index' of 46.

DISCRETE CASE

		STDPIN / DAY1			
		NONE	EQUIV	MOD	MARK
STDPIN / DAY2	NONE	258	31	3	0
	EQUIV		16	7	0
	MOD			10	2
	MARK				3

Table 7.16 Repeatability of the existing method when comparing the Inferior ST Depression Index (Day-to-day ECG recordings).

Using the maxima, minima and multiple steps techniques outlined in Chapter 5 it is possible to replace the ST 'score' with a smoothed alternative which should provide a greater resistance to fluctuations in any of the measurements which contribute to the score.

$$\begin{aligned} \text{STDPIN}_m^* &= -1\times\{\Pr[\min(x_1,x_2)<-1]+\Pr[\min(x_1,x_2)<-2]+ \\ &\quad +\Pr[\min(x_1,x_2)<-3]\} \\ &= -1\times\{(1-F_{11}\times F_{21})+(1-F_{12}\times F_{22})+(1-F_{13}\times F_{23})\} \end{aligned}$$

where $F_{ij}=\dfrac{e^{a_{ij}}}{1+e^{a_{ij}}}$; $a_{ij}=\dfrac{x_j-b_i}{0.05}$; b_i = threshold value $(i=1,2,3)$,

x_1 =ST Ampl. (lead II), x_2 = ST Ampl. (lead aVF).

STDPIN_m^* denotes the new smooth index of ST depression in the inferior leads.

Figure 7.3 illustrates the smooth representation of the Inferior ST depression index $STDPIN_m^*$ which has been calculated from the corresponding smooth indices of ST depression for leads II and aVF.

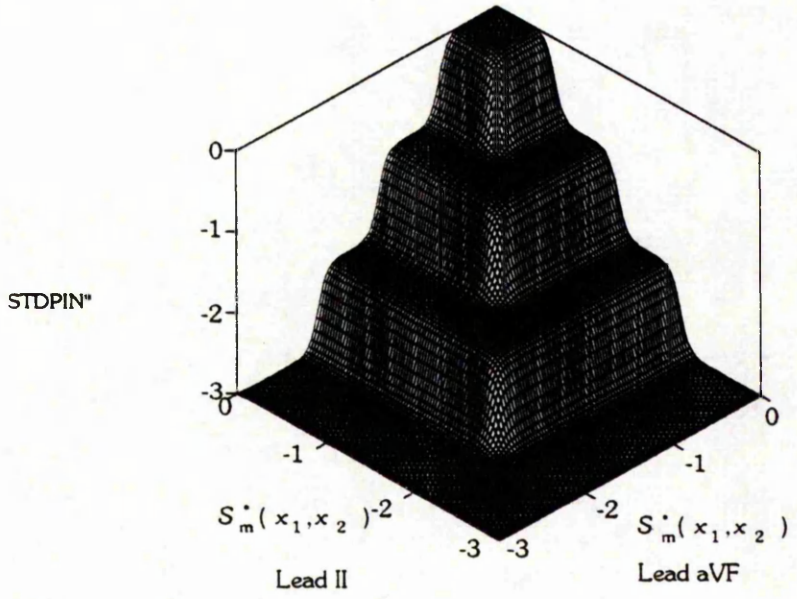


Figure 7.3 **Diagrammatic representation of smooth score function for Inferior ST Depression.**

Table 7.17 illustrates that the smoothing techniques have the advantage of eliminating many of the inconsistencies which have arisen using the established program. The 'Repeatability Index' has dropped from 46 to 25.

SMOOTHED CASE

		STDPIN / DAY1			
		NONE	EQUIV	MOD	MARK
STDPIN	NONE	282	15	1	0
DAY2	EQUIV		12	6	0
	MOD			9	2
	MARK				3

Table 7.17 **Repeatability of the smooth method when comparing the Inferior ST Depression Index (Day-to-day ECG recordings).**

The repeatability of the minute-to-minute ECG recordings was assessed using each method. Table 7.18 illustrates that there are 11 mild discrepancies and one of a severe nature when the existing method for identification of inferior ST depression is used. The corresponding 'Repeatability Index' is 14.

DISCRETE CASE

		STDPIN / MIN1			
		NONE	EQUIV	MOD	MARK
STDPIN	NONE	215	9	0	1
MIN2	EQUIV		15	2	0
	MOD			5	0
	MARK				2

Table 7.18 **Repeatability of the existing method when comparing the Inferior ST Depression Index (Minute-to-minute ECG recordings).**

Surprisingly, when the smooth method is adopted, the total numbers of mild and severe discrepancies do not change (although

there is some change in the pattern) and the 'Repeatability Index' remains at 14 (Table 7.19).

SMOOTHED CASE

		STDPIN / MIN1			
		NONE	EQUIV	MOD	MARK
STDPIN	NONE	224	6	0	1
MIN2	EQUIV		9	3	0
	MOD			3	2
	MARK				1

Table 7.19 **Repeatability of the smooth method when comparing the Inferior ST Depression Index (Minute-to-minute ECG recordings).**

It is also possible to assess the repeatability of each method using the split ECGs. Table 7.20 illustrates that the 'Repeatability Index' is 17 when the conventional method is used, and drops to 13 when the smooth techniques are applied (Table 7.21).

DISCRETE CASE

		ECG1			
		NONE	EQUIV	MOD	MARK
ECG 2	NONE	271	12	1	0
	EQUIV		26	3	0
	MOD			14	0
	MARK				3

Table 7.20 **Repeatability of the existing method when comparing the Inferior ST Depression Index (Split ECG recordings).**

SMOOTHED CASE

		ECG1			
		NONE	EQUIV	MOD	MARK
ECG 2	NONE	286	8	0	0
	EQUIV		18	5	0
	MOD			11	0
	MARK				2

Table 7.21 Repeatability of the smooth method when comparing the Inferior ST Depression Index (Split ECG recordings).

7.3.2 ST Elevation

Similarly, ST elevation is recorded if the maximum of the ST indices in either leads II or aVF is of a certain value although the diagnostic criteria is further complicated by the fact that an ST elevation index of 1 will only be recorded if the smaller of the ST indices in leads II and aVF is at least 1, i.e.

if $S(x_1, x_2) = 1$ in both leads then STELIN = 1

if $S(x_1, x_2) = 2$ in either lead then STELIN = 2

if $S(x_1, x_2) = 3$ in either lead then STELIN = 3

where STELIN denotes the index of ST elevation in the Inferior leads. Figure 7.4 shows this diagramatically.

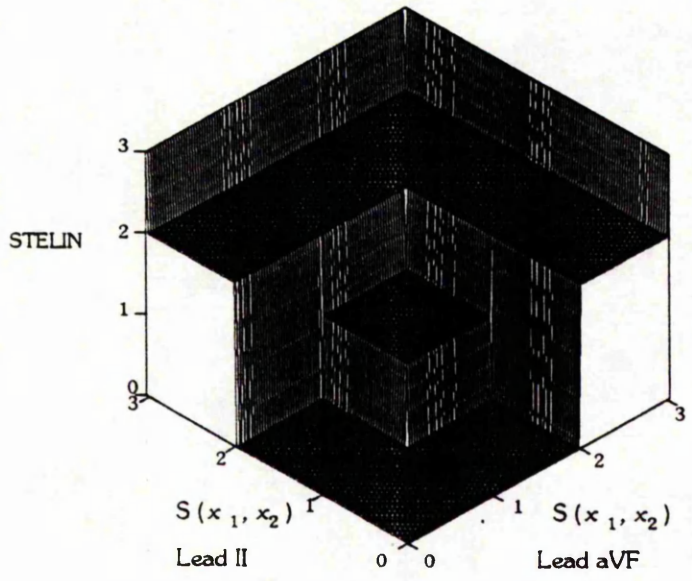


Figure 7.4 **Diagrammatic representation of discrete score function for Inferior ST Elevation.**

A smooth representation of STELIN may be calculated as follows:

$$\begin{aligned}\text{STELIN}_m^* &= 1\times\{\Pr[\min(x_1,x_2)>1 \text{ or } \max(x_1,x_2)>2]+\Pr[\max(x_1,x_2)>2]+\\ &\quad +\Pr[\max(x_1,x_2)>3]\}\\ &= 1\times\{F_{11}\times F_{21}+F_{12}\times F_{22}-F_{11}\times F_{21}\times F_{12}\times F_{22}+\\ &\quad F_{12}\times F_{22}+F_{13}\times F_{23}\}\end{aligned}$$

STELIN_m^* is the smooth index of ST elevation in the inferior leads (see Figure 7.5).

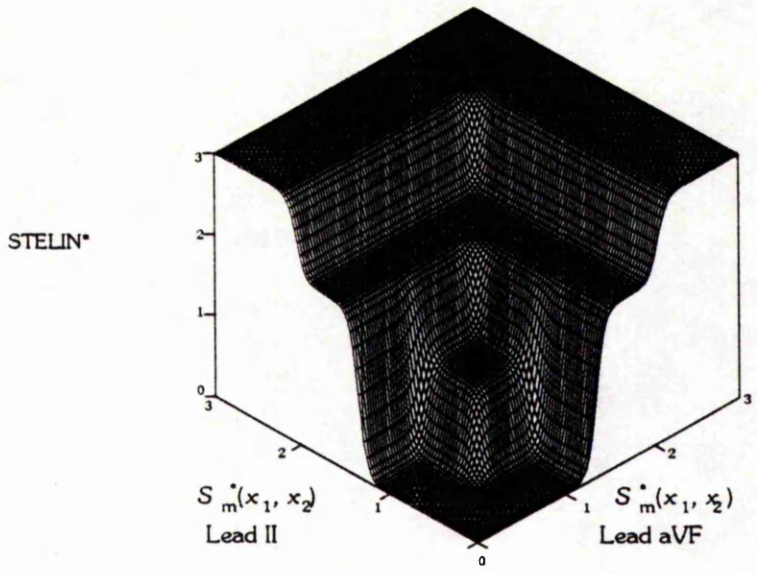


Figure 7.5 **Diagrammatic representation of smooth score function for Inferior ST Elevation.**

Comparisons of the indices of ST elevation for day-to-day, minute-to-minute and split ECGs may also be made using the methods which have been described. However, in the population studied, there were too few instances of ST elevation recorded for any realistic comparisons to be made. This is attributable to the fact that elevation of the ST segment is generally due to an acute condition not readily observed in our sample of patients who were selected because they were in a stable condition.

7.4 SUMMARY

In this chapter, many of the techniques that have previously been described were implemented in an attempt to improve the repeatability of the section of the Glasgow program which diagnoses abnormalities of the ST segment.

New smooth indices for ST depression were calculated for leads II and aVF and were subsequently combined to produce an overall index of ST depression for the inferior leads. These indices were then compared to the corresponding indices obtained when the conventional discrete method was used.

In terms of repeatability, the performance of the smooth method was considerably better than that of the discrete method when 330 day-to-day ECGs were assessed. The RI for the diagnosis of ST depression in lead II dropped from 35 to 19 (a reduction of 46%) when the discrete method was replaced by the smooth method. A 36% reduction in RI (from 28 to 18) was observed when comparing the repeatability of each method in identifying ST depression in lead aVF. The overall ST depression index for the inferior leads also benefited from an improvement in the level of agreement from day to day when the smooth method was used. The discrete method produced 40 mild and 3 moderate conflicting interpretations compared to only 23 and 1 respectively when the smooth method was used, representing a significant reduction in RI in the order of 46%.

Similar findings were reported when the repeatability of 249 minute-to-minute ECGs was assessed although the improvement gained was not as substantial. Although the repeatability of the smooth method was superior to that of the discrete method when the indices of ST depression in leads II and aVF were compared separately (with an 8% and a 50% improvement respectively), there was no noticeable improvement between the two methods when the overall index of ST depression in the inferior leads was considered.

One unexpected finding was that the repeatability of the smooth method was marginally worse than that of the discrete method when

the indices of ST depression in lead II from the split ECGs were compared. However, 3 of the mild discrepancies and the only moderate discrepancy observed in lead aVF when the split ECGs were considered were eliminated when the smooth version of the program was used. The smoothed version of the overall ST depression index for the inferior leads produced 13 minor discrepancies compared to 15 minor and one moderate discrepancy when the discrete version was used. This represents a 23% reduction in RI (from 17 to 13).

Proportionately fewer discrepant ECGs arise from the minute-to-minute and the split ECGs than from the day-to-day ECG recordings. For example, 32 of the 330 (9.7%) day-to-day ECGs produce inconsistent diagnoses for ST depression in lead II on the discrete method compared to 9 out of 249 (3.6%) for minute-to-minute ECGs and 9 out of 330 (2.7%) for split ECGs. Therefore there is less scope for improving on the 'Repeatability Index' for minute-to-minute and split ECGs which perhaps explains why it is more difficult to demonstrate any enhancement of repeatability consistently for the minute-to-minute and split ECGs.

CHAPTER EIGHT:

MYOCARDIAL INFARCTION.

8.1 INTRODUCTION.

A reduction in the coronary blood supply to the heart muscle can result in myocardial ischæmia, injury or infarction which can seriously impair the normal working process of the heart. A decrease in the flow of blood to the heart muscle can arise from inadequacies of the coronary circulation due to inflamed or blocked arteries and coronary spasm.

Myocardial ischæmia produces characteristic changes in the ECG in the form of inverted T waves and/or tall T waves in the appropriate lead or leads whereas myocardial injury is manifested in the ECG by elevation or depression of the ST segment depending on whether the injury is subepicardial or subendocardial. Necrotic areas of the heart are electrically silent so that in the presence of myocardial infarction electrodes placed over the infarcted area will record initially negative QRS complexes. The resultant of the electrical forces generated from the myocardium during ventricular depolarisation will tend to point away from the necrotic area producing abnormal Q waves in a lead overlying the area. The ECG is a useful tool in determining the extent and location of infarction and the detection of Q waves forms an integral part of the diagnostic process for myocardial infarction.

8.2 THE GLASGOW APPROACH.

The Glasgow Program diagnoses myocardial infarction in specific locations as a result of detecting Q waves with or without ST and T wave changes.

Pathological Q waves are identified if the amplitude and duration are of sufficient magnitude and/or the ratio of the Q wave amplitude to the R wave amplitude is large. Certain other criteria such as the magnitudes of the ST segment and the T wave are also considered in the diagnosis of myocardial infarction.

The repeatability of the section of the Glasgow program which is associated with the detection of myocardial infarction will be assessed by comparing the identification of inferior Q waves (i.e. pathological Q waves in any of the inferior leads II, III or aVF) in ECGs which have been recorded on two consecutive occasions.

8.3 IDENTIFICATION OF Q WAVES.

In the normal depolarisation process, the septum is depolarised from the left to the right side resulting in small Q waves being found in the lateral leads I, aVL, V5 and V6. It is necessary to be able to distinguish these insignificant findings from more substantial Q waves which would suggest that lateral myocardial infarction was present.

The Glasgow program uses a series of discrete criteria in order to identify whether there are pathological Q waves present in the various leads. This information is then used to make statements about the likelihood of myocardial infarction having taken place.

An example of a section of the diagnostic process for identifying the presence of Q waves on the basis of certain logical variables (Q1, Q2, Q3, Q4 etc.) is given in Table 8.1 below.

Q1	:	a)	i)	$Q_{DUR} > 0.035$secs **AND** $\lvert Q/R \rvert_{AMP} > 1/5$
		OR	ii)	$Q_{DUR} > 0.040$secs
	AND	b)		$Q_{AMP} < -0.09$mV **AND** peak-peak QRS > 0.15mV
Q2	:	a)	i)	$Q_{DUR} > 0.035$secs **AND** $\lvert Q/R \rvert_{AMP} > 1/5$
		OR	ii)	$Q_{DUR} > 0.030$secs **AND** $\lvert Q/R \rvert_{AMP} > 1/3$
	AND	b)		$Q_{AMP} < -0.20$mV
Q3	:	a)		$Q_{DUR} > 0.026$secs **AND** $\lvert Q/R \rvert_{AMP} > 1/5$
	AND	b)		$Q_{AMP} < -0.14$mV **AND** peak-peak QRS > 0.15mV
Q4	:	a)	i)	$Q_{DUR} > 0.020$secs **AND** $\lvert Q/R \rvert_{AMP} > 1/3$
		OR	ii)	$Q_{DUR} > 0.030$secs **AND** $T_{neg} < -0.10$mV
	AND	b)		$T_{neg} < -0.05$mV **OR** $ST_{AMP} > 0.06$mV
	AND	c)		$Q_{AMP} < -0.075$mV **AND** peak-peak QRS > 0.20mV

Table 8.1 **Diagnostic criteria in the Glasgow program for the identification of Q waves in the inferior and lateral leads (as of April 1990).**

The logical process outlined above is essentially graded in that Q1 defines the most pronounced Q waves whereas Q4 acknowledges shorter and shallower but nevertheless abnormal Q waves. Information is collected for all relevant leads and an overall decision as to whether pathological Q waves are present or absent in a particular set of leads is made. For example, abnormal Q waves in

the inferior leads (II, III and aVF) are identified if the following diagnostic statement is satisfied:

2 or more Q waves, including Q1 or Q2
OR A Q3 and 1(Q3 or QE3) all with $|Q/R| > 1/4$
OR One Q1 in II or aVF

(Note: In the above and Table 8.1, QE1 to QE4 are equivalent to Q1 to Q4 when Q waves are replaced by S waves and R amplitudes by R' amplitudes but with the proviso that the initial $r < 40\mu V$).

Again, owing to the nature of such criteria, there exists further potential for a lack of repeatability from one recording to the next. For example, if parts b) and c) of Q4 were satisfied and a Q wave of 0.019 secs in duration and -0.075mV in amplitude were recorded in any of the inferior leads, then none of the above conditions Q1 to Q4 would be satisfied whereas a Q wave of the same amplitude but 0.020 secs in duration would record a Q4.

8.4 SMOOTHING THE Q WAVE INDEX FOR THE INFERIOR LEADS.

One way of minimising repeat variation is to apply the techniques which have been introduced and used in earlier chapters. Discrete threshold values can be smoothed out by making use of the methods described in Chapter 5 so that Q1, Q2, Q3 and Q4 defined above as being present or absent may be represented by indices which can assess the degree to which a particular condition is satisfied.

The conditions outlined in Table 8.1 for Q1 can be rewritten in the following way:

$$Q1 = \begin{cases} 1 & \text{if } (Q_{DUR} > 0.035\text{secs } \textbf{AND } |Q/R| > 1/5 \\ & \textbf{OR } \quad Q_{DUR} > 0.040 \text{ secs}) \\ & \textbf{AND } Q_{AMP} < -0.09\text{mV } \textbf{AND} \\ & \qquad \text{peak-peak QRS} > 0.15\text{mV} \\ 0 & \text{otherwise} \end{cases}$$

Similar structures apply for Q2, Q3, Q4 etc.

The smoothed version of Q1 will then become

$$Q1^{*}_{s} = \{F_{11}F_{41} + F_{12} - F_{11}F_{41}F_{12}\} \times (1 - F_{21})F_{31}$$

where

$$F_{1i} = \frac{e^{a_{i_{x_1}}}}{1+e^{a_{i_{x_1}}}};\ a_{i_{x_1}} = \frac{x_1 - (a_i - \sigma_{x_1})}{\sigma_{x_1}};$$

$$F_{21} = \frac{e^{b_{x_2}}}{1+e^{b_{x_2}}};\ b_{x_2} = \frac{x_2 - (b - \sigma_{x_2})}{\sigma_{x_2}};$$

$$F_{31} = \frac{e^{c_{x_3}}}{1+e^{c_{x_3}}};\ c_{x_3} = \frac{x_3 - (c - \sigma_{x_3})}{\sigma_{x_3}};$$

$$F_{41} = \frac{e^{d_{x_4}}}{1+e^{d_{x_4}}};\ d_{x_4} = \frac{x_4 - (d - \sigma_{x_4})}{\sigma_{x_4}};$$

$$a_1 = 0.035\text{msecs};\ a_2 = 0.040\text{msecs};\ b = -0.09\text{mV}$$

$$c = 0.15\text{mV};\ d = \frac{1}{5}.$$

and x_1 denotes the Q wave duration,

x_2 denotes the Q wave amplitude,

x_3 denotes the peak-peak QRS,

x_4 denotes the Q/R ratio in the lead of interest

As before, σ_v denotes the estimated amount of day to day variation associated with ECG variable v. The appropriate Fortran coding can be seen in section C) of the Appendix.

Using rules of probability as analogues (see Chapter 5, section 5.5), a smoothed version which describes the extent to which Q

waves are present in the inferior leads can be calculated. Multi-parameter involvement coupled with the complexities of the diagnostic criteria and the requirement to consider three leads (II, III, aVF) instead of just one, results in a nontrivial equation which, for simplicity, will be split into three sections.

Thus

" 2 or more Q waves including Q1 or Q2 "

becomes

$$\Pr[A_1 \cap A_2] \cup \Pr[A_3 \cap A_4]$$

where A_1 is the event of a Q1 being present in either II, III or aVF,

A_2 is the event of at least one Q wave (excluding Q1) being present in leads II, III or aVF,

A_3 is the event of a Q2 being present in either II, III or aVF

and A_4 is the event of at least one Q wave (excluding Q2) being present in leads II, III or aVF.

Similarly,

" A Q3 and 1(Q3 or QE3) all with $|Q/R| > 1/4$ "

becomes

$$\Pr[B_1 \cap B_2 \cap B_3]$$

where

B_1 is the event that there is at least one Q3 in II, III, aVF,

B_2 is the event that there is at least one Q3 or QE3 in II, III or aVF

and B_3 is the event that $|Q/R| > 1/4$ in the lead where Q3 is true.

Lastly,

" Q1 in II or aVF "

becomes

$$\Pr[C_1 \cup C_2]$$

where

C_1 is the event that there is a Q1 in II and

C_2 is the event that there is a Q1 in aVF.

The new smooth Q wave index for the inferior leads will now range from 0 to 1 and will contain relevant information from all three inferior leads in a single estimate.

8.5 COMPARING THE REPEATABILITY OF THE EXISTING PROGRAM AND THE MODIFIED PROGRAM

8.5.1 Day-to-day ECG recordings.

Two ECGs were recorded on consecutive days for each of 330 inpatients in the Dept. of Medical Cardiology at the Royal Infirmary in Glasgow. These patients were considered to be in a stable condition. ECGs exhibiting evidence of conduction defects such as LBBB and RBBB were excluded from any analysis.

The discrete Q wave indices for the 330 day 1 and day 2 ECG recordings are tabulated in Table 8.2.

DISCRETE CASE

		DAY 1	
		NO Q WAVES	Q WAVES
DAY 2	NO Q WAVES	256	20
	Q WAVES		54

Table 8.2 **Repeatability of existing method. (Day-to-day ECG recordings)**

There are 20 pairs of consecutive ECGs which produce inconsistent diagnoses in terms of Inferior Q waves from day 1 to day 2 and the resulting Repeatability Index is 20.

In order to assess the repeatability of the smooth method, a cutpoint distinguishing significant and insignificant Q waves must be chosen.

A dotplot of the smoothed Q wave indices both for Q wave indices which have been assigned the value 0 using the existing discrete method and for those which have been assigned the value 1 is provided in Figure 8.1. These indices come from 330 day 1 ECG recordings and although we do not know if these discrete indices are correct, we make use of them in order to provide a means of obtaining a sensible cutpoint for the smooth representations.

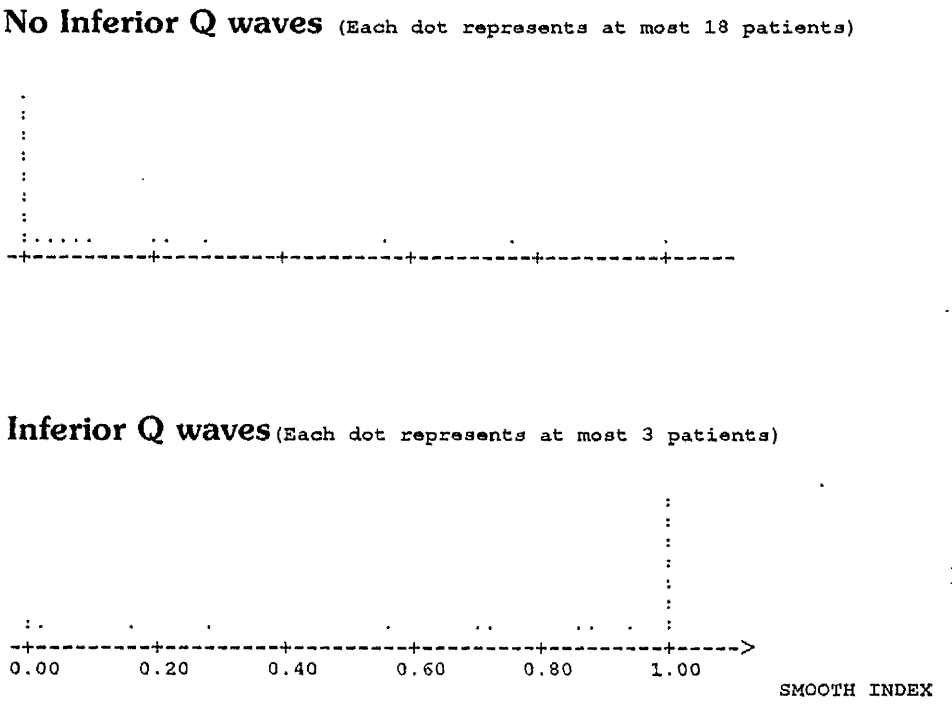


Figure 8.1 **Frequency Distribution of the smoothed Q wave index (Day 1 recordings).**

The cutpoint of 0.7 for separating those with inferior Q waves using the smooth method from those without produces the following table (Table 8.3) and has a repeatability index of 9.

SMOOTHED CASE

		DAY 1	
		NO Q WAVES	Q WAVES
DAY 2	NO Q WAVES	272	9
	Q WAVES		49

Table 8.3 **Repeatability of smoothed method. (Day-to-day ECG recordings)**

The day 1 and day 2 smoothed Q wave indices have been plotted against each other in Figure 8.2.

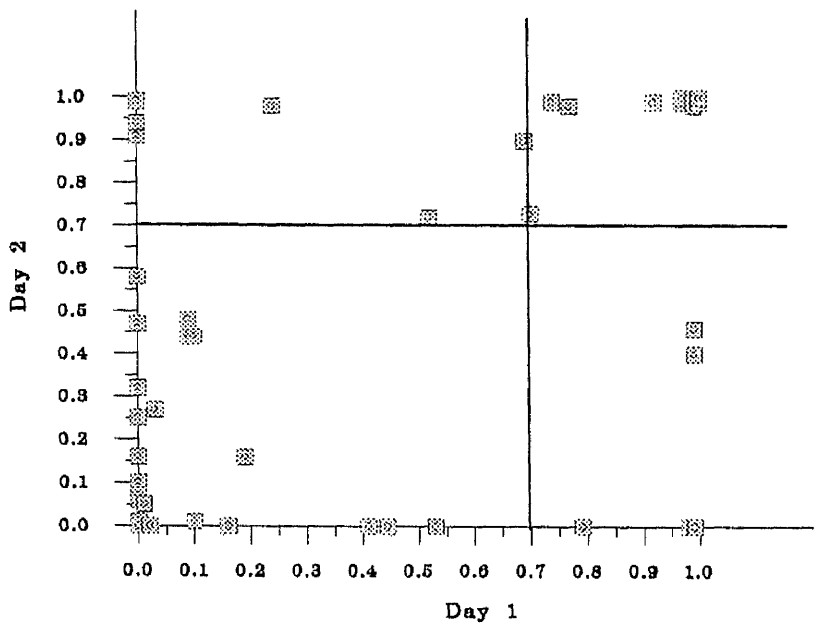


Figure 8.2 **Plot of Day 1 Q wave index vs. Day 2 Q wave index.**

From the plot it can be seen that some of the indices which are not in agreement from day 1 to day 2 lie close to the cutpoint of 0.7 while some are quite different from each other. However, there has

been a substantial improvement of around 55% in the repeatability of the program by adopting the smooth procedure.

8.5.2 Minute-to-minute ECG recordings.

The repeatability of the 249 pairs of ECGs which were recorded within the space of a few minutes was also assessed. Table 8.4 demonstrates that there were 9 inconsistent interpretations when the discrete method was used.

DISCRETE CASE

		MIN 1	
		NO Q WAVES	Q WAVES
MIN 2	NO Q WAVES	196	9
	Q WAVES		44

Table 8.4 **Repeatability of existing method. (Minute-Minute ECG recordings)**

Using the smoothing technique eliminates approximately half of this repeatability. Table 8.5 demonstrates that there are now only 5 cases of inconsistent reporting of Q waves from minute to minute. Compared to 9 cases using the conventional program, this represents an improvement of 44% in terms of the repeatability index.

SMOOTHED CASE

		MIN 1	
		NO Q WAVES	Q WAVES
MIN 2	NO Q WAVES	205	5
	Q WAVES		39

Table 8.5 **Repeatability of smoothed method. (Minute-Minute ECG recordings)**

8.5.3 Split ECG recordings.

Finally, the repeatability of the diagnosis of Q waves for both the existing and the smooth methods can be assessed in another way if we split the day 1 ECGs into two tracings as described earlier. Q wave indices using both the discrete and the smooth methods were obtained from the day 1 ECG tracings in this way. Table 8.6 demonstrates that there are 9 pairs of ECGs which produce inconsistent Q wave interpretations when the tracings are split in this way, compared to 6 when the smoothing techniques are adopted (Table 8.7).

DISCRETE CASE

		ECG 1	
		NO Q WAVES	Q WAVES
ECG 2	NO Q WAVES	261	9
	Q WAVES		60

Table 8.6 **Repeatability of existing method. (Split ECG recordings)**

SMOOTHED CASE

		ECG 1	
		NO Q WAVES	Q WAVES
ECG 2	NO Q WAVES	270	6
	Q WAVES		54

Table 8.7 **Repeatability of smoothed method. (Split ECG recordings)**

8.6 DISCUSSION.

This chapter has examined the effects that the smoothing techniques have on the detection of Q waves in the inferior leads. It must be stressed that these methods will also affect the subsequent diagnosis of inferior myocardial infarction which itself is suggested by the presence of Q waves. Unlike Chapter 7 where the identification of ST changes in the inferior leads may be buried within other diagnoses such as left ventricular hypertrophy and myocardial ischæmia, it is possible to assess the repeatability of inferior myocardial infarction statements which are produced by the diagnostic program developed in Glasgow.

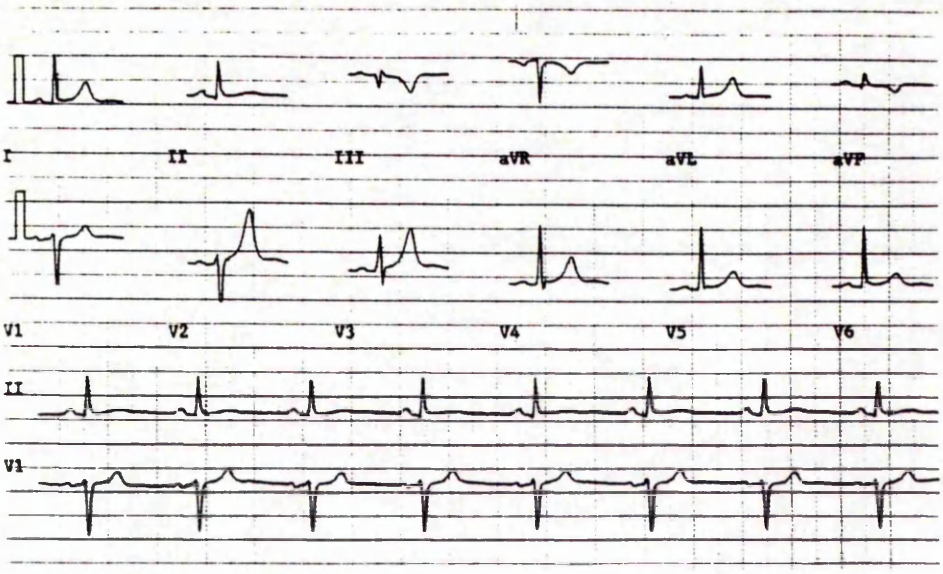
These established techniques may also be applied to the detection of Q waves in the anteroseptal, lateral and anterior leads and hence to the diagnosis of anteroseptal, lateral and anterior infarction. However, inferior Q waves were most prevalent in the database of day-to-day ECGs, providing a sufficient study group size.

As an additional exercise, an experienced electrocardiologist was given some pairs of ECGs from the database of day-to-day repeat ECGs to report. Some of the pairs were selected at random, although the ECGs which were particularly problematic with regard to day-to-day agreement were added to the final selection in order to assess the relative merits of the electrocardiologist and both the conventional and the modified versions of the Glasgow program. The day 1 ECGs were examined first, and the presence or absence of inferior Q waves (and hence inferior myocardial infarction) recorded. Several days elapsed before the corresponding day 2 ECGs were assessed to eliminate any possible learning effect.

Of the 34 pairs of ECGs which were examined in this way, 12 were given inconsistent diagnoses by the electrocardiologist compared to 19 and 9 by the conventional and the smoothed versions respectively. Since many of the pairs of ECGs for this small study were selected due to their initial incompatibility in terms of the identification of Q waves from day 1 to day 2 it was decided that further analysis was required to establish whether true changes in the ECG waveform did exist from day to day.

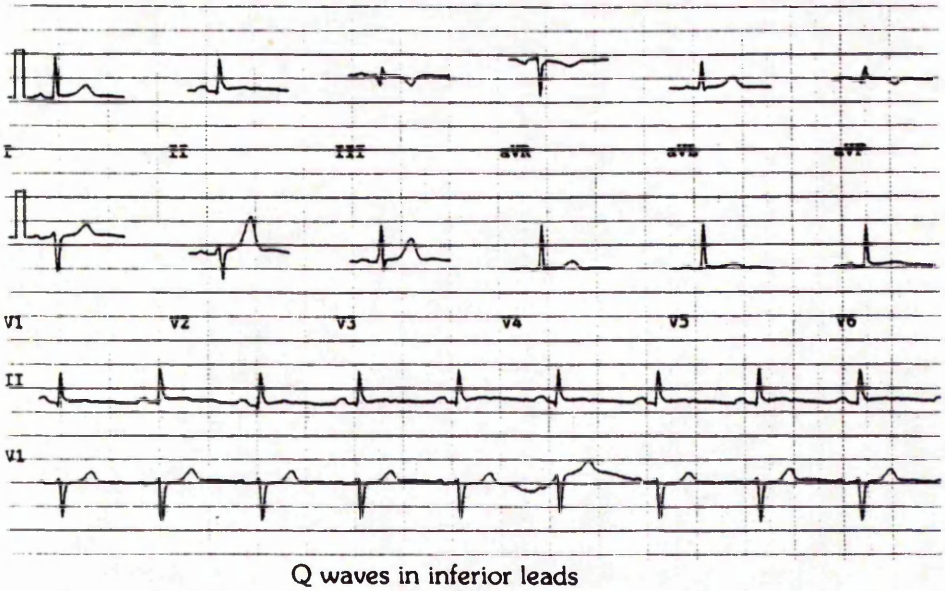
Instead of examining each of the ECGs separately, the reviewer scrutinised ECGs in pairs to look for any evidence of significant changes from one day to the next. His opinion was that 9 of the 34 pairs of ECGs did exhibit changes which could conceivably alter the diagnosis of inferior myocardial infarction. However, of these 9 pairs, only 4 were themselves a subset of the 12 incompatible diagnoses initially produced by the electrocardiologist when each ECG was examined individually. It was also discovered that 4 and 2 were subsets of the 9 and 19 discrepant ECG diagnoses produced by the smooth and the conventional methods respectively.

Figures 8.3 and 8.4 show 2 ECG recordings which were obtained from a 54 year old male patient on two consecutive days. On the second day (Figure 8.4), the conventional version of the diagnostic program reported evidence of inferior myocardial infarction whereas this statement was omitted for the initial recording (Figure 8.3). Both the modified version of the program and the experienced electrocardiologist reported inferior myocardial infarction on both occasions.



T wave changes in inferior leads
APPEARANCES ARE ABNORMAL AND MAY BE DUE TO MYOCARDIAL ISCHAEMIA

Figure 8.3 **Day 1 ECG for a 54 year old male together with the interpretation obtained with the conventional program.**

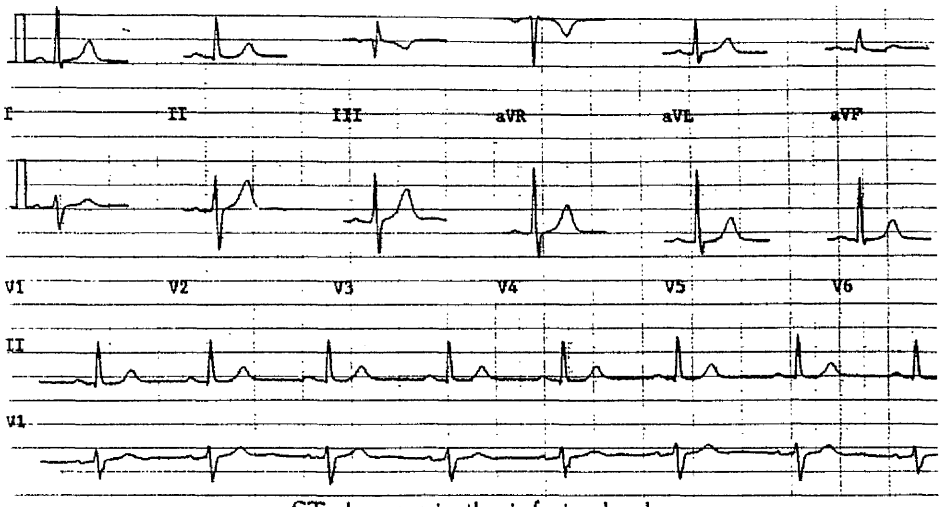


Q waves in inferior leads
T wave inversion also present
POSSIBLE INFERIOR INFARCTION - ? AGE

Figure 8.4 **Day 2 ECG for the same 54 year old male as in Fig 8.3 together with the interpretation obtained with the conventional program.**

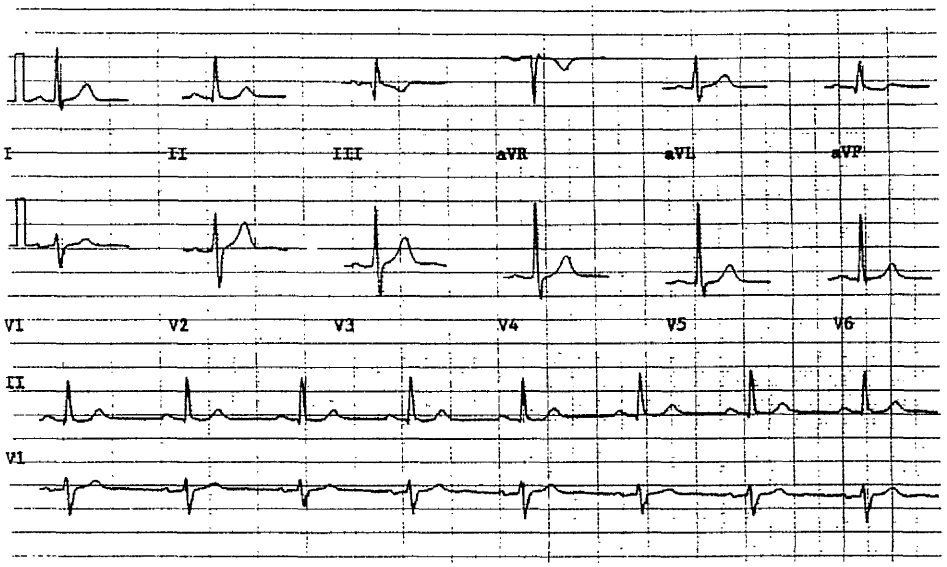
A similar situation can also be seen in Figures 8.5 and 8.6. The two ECGs have again been recorded on consecutive days, this time from a 44 year old male patient. While the human expert and the modified version of the program found no evidence of the presence

of an inferior infarct on either occasion, the conventional program reported inferior myocardial infarction on the second day (Figure 8.6).



ST changes in the inferior leads
THESE MINOR CHANGES ARE OF EQUIVOCAL SIGNIFICANCE ONLY

Figure 8.5 **Day 1 ECG for a 44 year old male together with the interpretation obtained with the conventional program.**



Q wave in inferior leads
T wave inversion also present
PROBABLE INFERIOR INFARCT - ? AGE

Figure 8.6 **Day 2 ECG for the same 44 year old male as in Fig 8.5 together with the interpretation obtained with the conventional program.**

In these two situations, the improvement in repeatability of the diagnostic program obtained when smoothing techniques have been implemented are in agreement with the findings of the electrocardiologist. However, this might not have been the case if the ECGs were examined independently by another clinician.

It should be noted that the repeatability of the determination of age of infarction was not investigated. This was not an issue of the current study although the methods which have been developed may be used to address this problem in the future. It was not feasible to obtain ECGs on line from acutely ill patients in the coronary care unit which is situated in the older part of Glasgow Royal Infirmary (GRI) and access was therefore limited to those in a stable condition in the cardiology wards in the new section of GRI. However, since ST and T wave changes are associated with the age and severity of infarction, the techniques outlined in Chapter 7 can be combined with the methods developed in the current chapter to assist in increasing the level of agreement between consecutive diagnoses of, for example, acute and/or old infarction in the future.

8.7 SUMMARY.

The repeatability of the section of the diagnostic program dedicated to the identification of inferior Q waves was investigated. The existing discrete method was compared to the smooth method and comparisons between day-to-day, minute-to-minute and split ECGs were made.

The smooth method produced more compatible diagnoses (with respect to Inferior Q waves) than the conventional method. Of the

330 day-to-day ECGs for which Q wave indices were provided, 20 exhibited inconsistencies when the existing method was used. This compared to only 9 inconsistencies (a reduction of around 55%) when the smooth method was used.

Of the 249 minute-to-minute ECGs which were assessed, there were 9 pairs exhibiting inconsistent diagnoses when the existing method was used, compared to only 5 when the smooth method was implemented. This represents an improvement of around 40% in terms of repeatability.

Lastly, the technique established by Bailey and his colleagues (Bailey et. al., 1974) allowed a further investigation of the repeatability of each method by splitting the day 1 ECGs which were initially sampled at 500 samples per second into odd and even samples at 250 samples/sec. Of the 330 day 1 ECG recordings which were split in this way, there were 9 inconsistent diagnoses using the existing method compared to 6 when the smooth method was used.

The diagnostic program currently in use in the Glasgow laboratory appears to be reasonably repeatable in terms of minute-to-minute and split ECG recordings. Although this means that it should be more difficult to improve the repeatability performance of the program, the new approach results in a reduction in the number of discrepant ECG interpretations.

In order to put the matter into perspective it is necessary to return to the results obtained by the reviewer of the repeated ECGs and to the corresponding results produced by both versions of the diagnostic program. Clearly, it is impossible to eliminate repeat variation completely. Human error will always prevail in the recording situation and in the interpretation of the ECG. In this attempt to

improve the repeated detection of inferior myocardial infarction using computer techniques it has become clear that although lack of repeatability may remain, its severity can be alleviated with the implementation of smoothing techniques.

CHAPTER NINE: CONCLUSIONS.

The principal objective of this study has been to develop a methodology for improving the repeatability of computer assisted interpretation of electrocardiograms. The following important conclusions can be drawn.

9.1 OVERALL REPEATABILITY.

In the various individual chapters, results were presented which showed that a definite improvement in the repeatability of certain individual diagnoses could be obtained. In other words, computer assisted ECG interpretation can be enhanced by the methods developed in this study.

It is of interest to consider the cumulative effect of improving the diagnosis of left ventricular hypertrophy (LVH), ST depression in the inferior leads and inferior myocardial infarction (IMI). Overall repeatability of the Glasgow program was assessed on the basis of the level of agreement between consecutive ECG recordings obtained

a) one day apart,
b) within the space of a few minutes, and
c) using a method of artificially splitting one ECG into two digital representations

with respect to the presence or absence of the three previously mentioned conditions. Of the 330 pairs of ECGs which were

recorded one day apart, 266 were in total agreement for all the conditions mentioned when the conventional version of the program was used, representing an overall repeatability of 81%. In comparison, the repeatability of the modified version of the program (using the same set of day-to-day ECGs) was 88% (i.e. 291 of the 330 pairs of ECGs produced consistent diagnoses).

Due to the elimination of various external sources of variation such as electrode positioning and recording technique, it was expected that the ECGs which were recorded within the space of several minutes and the ECGs which were split into two digital representations would demonstrate a higher level of agreement. Indeed, this was the case, with 222 of the 249 (89%) pairs of minute-to-minute ECGs and 304 of the 330 (92%) pairs of split ECG recordings producing consistent diagnoses using the conventional method. However, smoothing techniques continue to enhance the repeatability of both sets of ECGs, with both the percentages of minute-to-minute ECGs and split ECGs demonstrating agreement in the three areas mentioned of approximately 95%, representing overall reductions in error of 55% and 38% respectively.

9.2 ENHANCED PERFORMANCE.

The level of agreement from ECG to ECG is significantly superior when the smoothing methods are used. When considering the ECGs which were repeated on consecutive days it can be shown that the approximate 95% confidence interval for the difference in the proportion of pairs of ECGs which are not in agreement from day to day is (4%, 12%). This means that, at worst, the modified

version of the Glasgow program is 4% more repeatable and at best 12% more repeatable than the conventional program. The corresponding confidence intervals for the ECGs which were repeated from one minute to the next and for the ECGs which were split into odd and even samples are (1%, 10%) and (0%, 7%) which demonstrate that in general the smoothing techniques are having the desired effect of enhancing the repeatability of the Glasgow program.

9.3 COMPARATIVE RESULTS.

It is of interest to contrast the degree of consistency of the modified version of the Glasgow program with the levels of agreement obtained in diagnostic statements from previous studies.

Having established an artificial method of enabling developers and users of a particular automated system to assess its repeatability, Bailey et. al. (1974) used this technique on various programs. Reproducibility of the ECG interpretations was poor, with the percentages of 'identical readings' ranging from 49.8% for version D of the PHS program to 82.4% for the AVA (3.4) program. When these results are compared to the repeatability of 92% for the conventional Glasgow program and 95% for the modified version, it is clear that improvements have been obtained.

Machado et. al. (1991) investigated the minute-to-minute repeatability of the 1988 release of the unmodified Glasgow program on a sample of 410 pairs of ECGs which were recorded, without removal of the electrodes, using commercially available equipment. Identical type A statements, i.e. those that can be verified using non-

ECG methods, were produced in 93% of cases. Of the 249 pairs of minute-to-minute ECGs examined in the current study, repeatability was also of the order of 93% for the conventional Glasgow program when considering Type A statements relating to LVH and IMI only. The modified version of the Glasgow program displayed an improved (and near perfect) repeatability of approximately 98% and would almost certainly have had the effect of increasing the reproducibility of the type A statements in the 410 patients previously investigated by Machado.

Method	**24 hrs apart**		**1 min apart**		**Artificially split**	
n	660{330 Day 1, 330 Day 2}		498{249 Min 1, 249 Min 2}		660{330 odd samples, 330 even samples}	
Repeat-ability(%)	**OLD**	**NEW**	**OLD**	**NEW**	**OLD**	**NEW**
Overall	81	88	89	95	92	95
Type A	89	95	93	98	96	98

Table 9.1 **Summary of results**

A brief summary of results can be seen in Table 9.1. Throughout, the percentage of ECGs which were in agreement from day to day was lower than that observed in ECGs which were recorded in the space of several minutes and in ECG recordings which were artificially split into two. This was to be expected given that there was more potential for external sources of variation, such as the removal and replacement of electrodes, to contribute to lack of repeatability. The repeatability of statements relating to Type A conditions, in this case LVH and IMI only, was greater in all three

cases than that observed when ST changes were also considered. This would suggest that the agreement in diagnoses relating to ST abnormalities is a particularly problematic area requiring further research.

9.4 FUTURE DEVELOPMENTS.

There is still scope for further developments in the repeatability of computer assisted interpretation of the electrocardiogram.

This study has dealt with only three diagnostic categories of the Glasgow ECG Analysis program and has not, for example, considered the repeatability in the diagnosis of anterior myocardial infarction. Also, many of the 5% of pairs of minute-to-minute ECGs which demonstrated inconsistencies did so as a result of LVH scores, Q wave indices or indices of Inferior ST depression lying close to the boundary values. Increasing the size of the database of ECGs over a period of time may result in alternative cutpoints being suggested which could have the effect of improving repeatability even more. However, it is important to note that the selection of optimal cutpoints will always remain problematic, with training datasets suggesting values which may not work as effectively in subsequent test sets.

The computer assisted interpretation of the ECG is clearly an invaluable diagnostic tool. However, this study has demonstrated that the technique is not perfect, especially in terms of repeatability. Methods have been developed in order to address the problem of lack of repeatability and the results are encouraging. It only remains to point out that similar techniques ought to be applied to other areas of

the Glasgow program and it is likely that overall repeatability will be further enhanced.

REFERENCES

Bailey, J. J., Itscoitz, S. B., Hirshfield, J. W., Grauer, L. E., Horton, M. (1974).
A method for evaluating computer programs for electrocardiographic interpretation.
I Application to the experimental IBM program of 1971.
II Application to version D of the PHS program and the Mayo Clinic program of 1968.
III Reproducibility testing and the sources of program errors.
Circulation; I: **50**; 73-79, II: **50**; 80-87, III: **50**; 88-93.

Bailey, J. J., Horton, M., Itscoitz, S. B. (1976).
The importance of reproducibility testing of computer programs for electrocardiographic interpretation: Application to the Automatic Vectorcardiographic Analysis program (AVA 3.4).
Computers and Biomedical Research; **9**: 307-316.

Bailey, J. J., Horton, M. R. (1977).
Advantages of automation of ECG analysis with conventional (heuristic) criteria.
In: Trends in Computer-Processed Electrocardiograms: ed. van Bemmel, J. H., Willems, J. L., pp 221-228. Amsterdam, North Holland.

Bailey, J. J., Campbell, G., Horton, M. R., Shrager, R. I., Willems, J. L. (1988).
Determination of statistically significant differences in the performance of ECG Diagnostic Algorithms: An improved method.
Journal of Electrocardiography; **21**: (Suppl.) S188-S192.

Barnett, V., Lewis, T. (1978).
Outliers in Statistical Data.
Chichester, Wiley.

Bonner, R. E., Crevasse, L., Ferrer, M. I., Greenfield, J. C. (1972).
A new computer program for analysis of scalar electrocardiograms.
Computers and Biomedical Research; **5**: 629-653.

Brohet, C. R., Robert, A., Derwael, C., Fesler, R., Stijns, M., Vliers, A., Braasseur, L. A. (1984).
Computer interpretation of pediatric orthogonal electrocardiograms: statistical and deterministic classification methods.
Circulation; **70**: 255-263.

Burdon-Sanderson, J., Page, F. J. M. (1878).
Experimental results relating to the rhythmical and excitatory motions of the ventricle of the heart of the frog and of the electrical phenomena which accompany them.
Proceedings of the Royal Society of London; **7**: 410-414.

Burger, H.C. and van Milaan, J. B. (1946), (1947), (1948).
Heart-vector and leads. I, II and III.
British Heart Journal;(1946) **8**: 157-161, (1947) 9: 154-160, (1948)10: 229-233.

Caceres, C. A., Steinberg, C. A., Abraham, S., Carberry, W. J., McBride, J. M., Tolles, W. E., Rikli, E. D. (1962).
Computer extraction of electrocardiographic parameters.
Circulation; **25**: 356-362.

Caceres, C. A. (1963).
Electrocardiographic analysis by a computer system.
Archives of Internal Medicine; **111**: 114-120.

Cady, L. D., Woodbury, M. A., Tick, L. J., Gertler, M. M. (1961).
A method for electrocardiogram wave-pattern estimationExample : Left ventricular hypertrophy.
Circulation Research; **9**: 1078-1082.

Cawood, H. T., Hillis, S., Macfarlane, P. W., Lawrie, T. D. V. (1974).
Day to day and beat to beat variation in normal 3 orthogonal lead electrocardiograms.
British Heart Journal; **36**: 1225-1230.

Chen, C. Y., Chiang, B. N., Macfarlane, P. W. (1989).
Normal limits of the electrocardiogram in a Chinese population.
Journal of Electrocardiology; **22**: 1-15.

Committee of the American Heart Association for the standardisation of precordial leads.
Standardisation of Precordial Leads.
(a) Supplementary report (1938).
American Heart Journal; **15**: 235-239.
(b) Second Supplementary Report (1943).
Journal of the American Medical Association; **121**: 1349-1351.

Cornfield, J., Dunn, R. A., Batchlor, C. D., Pipberger, P. V. (1973).
Multigroup diagnosis of electrocardiograms.
Computers and Biomedical Research; **6**: 97-120.

Degani, R., Bortolan, G. (1986).
Combining measurement precision and fuzzy diagnostic criteria.
In: Computer ECG analysis: Towards standardisation, ed. Willems, J. L., van Bemmel, J. H., Zywietz, C. pp 177-182. Amsterdam, North Holland.

Eddelman, E. E. and Pipberger, H. V. (1971).
Computer analysis of the orthogonal electrocardiogram and vectorcardiogram in 1,002 patients with myocardial infarction.
American Heart Journal; **81**: 608-621.

Einthoven, W. (1901).
Un nouveau galvonomètre.
Arch. Neerl. Sci. Exactes Nat.; **6**: 625-633.

Einthoven, W., Fahr, G., de Waart, A. (1913).
Uber die Richtung und die manifeste Gross der potentialschwankungen im menschlichen Herzen and uber den Einfluss der Herzlage auf die form des Elektrokardiogramms.
Pleugers Arch; **150**: 275-315.
(Translation: Hoff, H. E., Sekelj, P. American Heart Journal (1950); **40**: 163-94.)

Frank, E. (1956).
An accurate, clinically practical system for spatial vectorcardiography.
Circulation; **13**: 737-749.

Goldberger, E. (1942).
A simple, indifferent, electrocardiographic electrode of zero potential and a technique of obtaining augmented unipolar extremity leads.
American Heart Journal; **23**: 483-492.

Goldman, M. J. and Pipberger, H. V. (1969).
Analysis of the orthogonal electrocardiogram and vectorcardiogram in ventricular conduction defects with and without myocardial infarction.
Circulation; **34**: 243-250.

Graybiel, A., Starr, R. S., White, P. D. (1938).
Electrocardiographic changes following the inhalation of tobacco smoke.
American Heart Journal; **15**: 89-99.

Gubner, R., Ungerleider, H. E. (1943).
Electrocardiographic criteria of left ventricular hypertrophy.
Archives of Internal Medicine; **72**: 196-209.

Hampel, F. R. (1974).
The Influence Curve and its Role in Robust Estimation.
Journal of the American Statistical Association; **69**: 383-393.

Hermann, G. R., Wilson, F. N. (1922).
Ventricular Hypertrophy, comparison of ECG and post-mortem observations.
Heart; **9**: 91-98.

Huwez, F. U. (1990).
Electrocardiography of the Left Ventricle in Coronary Artery Disease and Hypertrophy.
Ph.D. Thesis: University of Glasgow.

Kalbfleisch, J. G. (1979).
Probability and Statistical Inference II. New York, Springer-Verlag.

Kimura, E., Mibukura, Y., Miura, S. (1963).
Statistical diagnosis of electrocardiogram by Theorem of Bayes. A preliminary report.
Japanese Heart Journal; **4**: 469-488.

Klingeman, J. and Pipberger, H. V. (1967).
Computer classifications of electrocardiograms.
Computers and Biomedical Research; **1**: 1-17.

Kors, J. A., van Bemmel, J. H. (1990).
Classification methods for computerised interpretation of the electrocardiogram.
Methods of Information in Medicine; **29**: 330-336.

Kossmann, C. E., Johnston, F. D. (1935).
The precordial electrocardiogram I. The potential variations of the precordium and of the extremities in normal subjects.
American Heart Journal; **10**: 925-941.

Lewis, T., Gilder, M.D. (1912).
The Human Electrocardiogram: A preliminary investigation of young male adults to form a basis for pathological study.
Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences (London); **202**: 351-376

Lewis, T. (1914).
Observations upon ventricular hypertrophy, with special reference to preponderance of one or other chamber.
Heart; **5**: 367-402.

Macfarlane, P. W. (1971).
ECG waveform identification by digital computer.
Cardiovascular Research; **5**:141-146.

Macfarlane, P. W., Lawrie, T. D. V. (1974).
An Introduction to Automated Electrocardiogram Interpretation.
London, Butterworths.

Macfarlane, P. W., Cawood, H. T., Lawrie, T. D. V. (1975).
A basis for computer interpretation of serial electrocardiograms.
Computers and Biomedical Research; **8**: 189-200.

Macfarlane, P.W., Watts, M. P., Lawrie, T. D. V. (1980).
In: Optimisation of Computer ECG Processing: ed. Wolf, H. K., Macfarlane, P.W. Hybrid Electrocardiography. pp 57-61.

Macfarlane, P.W., MacLean, J. A., Low, M. et al. (1985).
Revised limits of the normal ECG.
In: Advances in Electrocardiology. ed. D'Alché, P., pp. 160-163.
Caen: Centre de Publications de l'Université de Caen.

Macfarlane, P. W. (1989a).
In: Comprehensive Electrocardiology. ed. Macfarlane, P. W., and Lawrie, T. D. V. The Coming Age of Electrocardiology, pp 3-40. Oxford, Pergamon Press.

Macfarlane, P. W. (1989b).
In: Comprehensive Electrocardiology. ed. Macfarlane, P. W., and Lawrie, T. D. V. Lead Systems, pp. 315-352. Oxford, Pergamon Press.

Macfarlane, P. W. (1989c).
The Sicard 440/ Sicard 740 ECG analysis program.
Electromedica; **57**: 106-120.

Macfarlane, P. W., Lawrie, T. D. V. (1989d).
In: Comprehensive Electrocardiology. ed. Macfarlane, P. W., and Lawrie, T. D. V. The Normal Electrocardiogram and Vectorcardiogram, pp. 407-457. Oxford, Pergamon Press.

Macfarlane, P.W, Lawrie, T. D. V. (1989e).
In: Comprehensive Electrocardiology. ed. Macfarlane, P. W. and Lawrie, T. D. V. Appendix 1: Normal Limits, pp. 1141-1526. Oxford, Pergamon Press.

Macfarlane, P. W., Devine, B., Latif, S., McLaughlin, S. C., Shoat, D. B., Watts, M. P. (1990).
Methodology of ECG interpretation in the Glasgow program.
Methods of Information in Medicine; **29**: 354-361.

Machado, H. B., Emilia, M., Silva, P., Lourenco, L. (1991).
Reproducibility of diagnostic interpretative statements in a leading ECG computerised system.
Proceedings of Melecon '91; **1**: 793-795.

Marey, E. J. (1876).
Des variations électriques des muscles et du coeur en particulier étudies au moyen de l'électromètre de M. Lippmann.
Comptes Rendus de l'Academie des Sciences; **82**: 975-977.

McNemar, Q. (1947).
Note on the sampling error of the difference between correlated proportions or percentages.
Psychometrika; **12**: 153-157.

Milliken, J. A., Macfarlane, P. W., Lawrie, T. D. V. (1989).
In: Comprehensive Electrocardiology. ed. Macfarlane, P. W., and Lawrie, T. D. V. Enlargement and Hypertrophy, pp. 631-670. Oxford, Pergamon Press.

Noth, P. H., Myers, G. B., Klein, H. A. (1947).
The precordial electrocardiogram in LVH: A study of autopsy cases.
Journal of Laboratory and Clinical Medicine; **32**: 1517.

Pipberger, H. V., Freis, E. D., Taback, L., Mason, H. L. (1960).
Preparation of ELectrocardiographic Data for Analysis by Digital Electronic Computer.
Circulation; **21**: 413-418.

Pipberger, H. V., McCaughan, D., Littmann, D., Pipberger, H. A., Cornfield, J., Dunn, R. A., Batchlor, C. D., Berson, A. S. (1975).
Clinical Application of a Second Generation Electrocardiographic Computer Program.
American Journal of Cardiology; **35**: 597-608.

Romhilt, D. W., Estes, E. H. (1968).
A point scoring system for the electrocardiographic diagnosis of left ventricular hypertrophy.
American Heart Journal; **75**: 752-758.

Sears, G. A., Manning, G. W. (1958).
Routine Electrocardiography: Postprandial T-Wave Changes.
American Heart Journal; **56**: 591-597.

Silverman, B. W. (1985).
Some aspects of the spline smoothing approach to non-parametric regression curve fitting.
Journal of the Royal Statistical Society B; **47**: 1-52.

Silverman, B. W. (1986).
Density Estimation for statistics and data analysis. London, Chapman and Hall.

Simonson, E., Brozek, J., Keys, A. (1949).
Variability of the electrocardiogram in normal young men.
American Heart Journal; **38**: 407-422.

Simonson, E. (1961).
Differentiation between normal and abnormal in Electrocardiography.
St. Louis, The C. V. Mosby Company.

Smets, Ph., Kornreich, F., Block, P., Bernard, R., Vainsel, H. (1977).
Fuzzy diagnostic, degree of belief and utility.
In: Trends in Computer-Processed Electrocardiograms, ed. van Bemmel, J. H., Willems, J. L., pp 257-259. Amsterdam, North Holland.

Sokolow, M., Lyon, T. P. (1949).
The ventricular complex in left ventricular hypertrophy as obtained by unipolar precordial and limb leads.
American Heart Journal; **67**: 161-186.

Stallmann, F. W., Pipberger, H. V. (1961).
Automatic recognition of electrocardiographic waves by a digital computer.
Circulation Research; **9**: 1138-1143.

Tuinstra, C. L. (1986).
Reproducibility of ECG interpretation statements.
In: Computer ECG analysis: Towards standardisation, ed. Willems, J. L., van Bemmel, J. H., Zywietz, C. pp 341-346. Amsterdam, North Holland.

van Bemmel, J. H. (1986).
Past and future research goals for computerised ECG processing.
In: Computer ECG Analysis: Towards standardisation, ed. Willems, J. L., van Bemmel, J. H., Zywietz, C. pp 367-381. Amsterdam, North Holland.

van Bemmel, J. H., Duisterhout, J. S., van Herpen, G., Bierwolf, L. G., Hengeveld, S. J., Versteeg, B. (1971).
Statistical processing methods for recognition and classification of vectorcardiograms.
In: Vectorcardiography 2, ed. Hoffman, I., Hamby, R. I., Glassman, E., Proc. 11th International Vectorcardiography Symposium; pp 207-215. Amsterdam, North Holland.

Waller, A. D. (1887).
A demonstration on man of electromotive changes accompanying the heart's beat.
Journal of Physiology; **8**: 229-234.

Waller, A. D. (1889).
On the electromotive changes connected with the beat of the mammalian heart, and of the human heart in particular.
Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences (London); **180**: 169-194.

Walston, A., Harley, A., Pipberger, H. V. (1974).
Computer Analysis of the orthogonal electrocardiogram and vectorcardiogram in mitral stenosis.
Circulation; **50**: 472-477.

Watts, M. P., Shoat, D. B. (1987).
Trends in electrocardiographic design.
Journal of the Institution of Electronic and Radio Engineers; **57**: 140-150.

Willems, J. L., Poblete, P. F., Pipberger, H. V. (1972).
Day-to-day variation of the normal orthogonal electrocardiogram and vectorcardiogram.
Circulation; **45**: 1057-1064.

Willems, J. L. (1977).
Introduction to multivariate and conventional computer ECG analysis: Pro's and contra's.
In: Trends in Computer-Processed Electrocardiograms, ed. van Bemmel, J. H., Willems, J. L., pp 213-220. Amsterdam, North Holland.

Willems, J. L., Lesaffre, E., Pardaens, J., de Schreye, D. (1986).
Multigroup logistic classification of the standard 12- and 3-lead ECG.
In: Computer ECG Analysis: Towards standardisation, ed. Willems, J. L., van Bemmel, J. H., Zywietz, C., pp 203-210. Amsterdam, North Holland.

Willems, J. L., Abreu-Lima, C., Arnaud, P., van Bemmel, J. H., Brohet, C., Degani, R., Denis, B., Graham, I., van Herpen, G., Macfarlane, P. W., Michaelis, J., Moulopoulos, S., Poppl, S., Zywietz, C. (1987).
Testing the performance of ECG computer programs: The CSE diagnostic pilot study.
Journal of Electrocardiography;**20**: (Suppl.); S73-S77.

Willems, J. L. (1988).
Assessment of the performance of ECG computer programs: The CSE diagnostic pilot study.
IEEE: 41-46.

Willems, J. L., Abreu-Lima, P., Arnaud, P., Brohet, C. R., Denis, B., Gehring, J., Graham, I., van Herpen, G., Machado, H., Michaelis, J., Moulopoulos, S. D. (1990).
Evaluation of ECG interpretation results obtained by computer and cardiologists.
Methods of Information in Medicine; **29**: 308-316.

Willems, J. L., Abreu-Lima, C., Arnaud, P., van Bemmel, J. H., Brohet, C., Degani, R., Denis, B., Gehring, J., Graham, I., van Herpen, G., Machado, H., Macfarlane, P. W., Michaelis, J., Moulopoulos, S. D., Rubel, P., Zywietz, C. (1991).
The Diagnostic performance of computer programs for the interpretation of electrocardiograms.
New England Journal of Medicine; **352**: 1767-1773.

Williams, H. B. (1914).
On the cause of the phase difference frequently observed between homonymous peaks of the electrocardiogram.
American Journal of Physiology; **35**: 292-300.

Wilson, F. N., Johnston, F. D., Garrard, A., Macleod, M. D., Barker, P. S. (1934).
Electrocardiograms that represent the potential variations of a single electrode.
American Heart Journal; **9**: 447-471.

Wilson, F. N., Johnston, F. D., Rosenbaum, F. F., Erlanger, H., Kossmann, C. E., Hecht, H., Cotrim, N., Menezes de Oliveira, R., Scarsi, R., Barker, P. S. (1944).
The precordial electrocardiogram.
American Heart Journal; **27**: 19-85.

APPENDIX

In order to implement the smoothing techniques which have been developed in this thesis, it was necessary to re-write sections of the diagnostic Fortran code. This task required several thousand new lines of code and the following small examples serve as illustrations for the three sections of the program which have already been adapted.

A) The following section of Fortran code calculates the contribution to the new smooth LVH index made by the amplitude of the R wave in lead V5 or lead V6.

```
      est1d=(la(11).and.la(12))
      rmaxv5=max(iabs(meas(jra,11)),iabs(meas(jrda,11)))
      rmaxv6=max(iabs(meas(jra,12)),iabs(meas(jrda,12)))
      sd5=sqrt(1024.0+(13.8*rmaxv5))
      sd6=sqrt(0.64+(13.8*rmaxv6))
c**** set up the continuous limits for RV5/RV6 ****
      lvhlim(1,agep,4,1)=(59.77-0.01089*age)**2
      lvhlim(2,agep,4,1)=(47.654-0.00273*age)**2
      lvhlim(1,agep,4,2)=(47.664-0.000315*age)**2
      lvhlim(2,agep,4,2)=(37.115-0.00904*age)**2
      if(.not.est1d.or.agep.le.5) goto 140
      vara=exp((rmaxv5-(lvhlim(sexp,agep,4,racep)-sd5))/sd5)
      varb=exp((rmaxv5-(lvhlim(sexp,agep,4,racep)+500-sd5))/sd5)
      varc=exp((rmaxv5-(lvhlim(sexp,agep,4,racep)+1000-sd5))/sd5)
      vard=exp((rmaxv5-(lvhlim(sexp,agep,4,racep)+1500-sd5))/sd5)
      vare=exp((rmaxv6-(lvhlim(sexp,agep,4,racep)-sd6))/sd6)
      varf=exp((rmaxv6-(lvhlim(sexp,agep,4,racep)+500-sd6))/sd6)
      varg=exp((rmaxv6-(lvhlim(sexp,agep,4,racep)+1000-sd6))/sd6)
      varh=exp((rmaxv6-(lvhlim(sexp,agep,4,racep)+1500-sd6))/sd6)
```

```
c****  calculate the new smooth idscore ****

305    idscor1=2*(vara/(1+vara))+(varb/(1+varb))+(varc/(1+varc))+
     .                (vard/(1+vard))
306    idscor2=2*(vare/(1+vare))+(varf/(1+varf))+(varg/(1+varg))+
     .                (varh/(1+varh))
```

B) The relevant Fortran coding for the calculation of the new smoothed ST index can be seen. This new ST index is subsequently used in the identification of ST changes.

```
c**** initialise and calculate ST Index for I,II,III,AVL,AVF,V5,V6 ****
      do 10 i=1,7
        stscor(i) = 0
        varv1(i) = 1
        varv2(i) = sqrt(sig(i)+taw(i)*abs(meas(jsta,l(i))))
        vara(i) = exp((-100+varv2(i)-meas(jsta,l(i)))/varv2(i))
        varb(i) = exp((-50+varv2(i)-meas(jsta,l(i)))/varv2(i))
        varc(i) = exp((0+varv2(i)-meas(jsta,l(i)))/varv2(i))
        vard(i) = exp((0-meas(jsts,l(i)))/varv1(i))
        vare(i) = exp((meas(jsta,l(i))-(60-varv2(i)))/varv2(i))
        varf(i) = exp((meas(jsta,l(i))-(80-varv2(i)))/varv2(i))
        varg(i) = exp((meas(jsta,l(i))-(100-varv2(i)))/varv2(i))
        varh(i) = exp((meas(jsts,l(i))-0)/varv1(i))
        vari(i)=3*(exp((meas(jsta,l(i))-(100-varv2(i)))/varv2(i))
        part1(i) = -((vara(i)/(1+vara(i)))+(varb(i)/(1+varb(i)))+
     .                  (varc(i)/(1+varc(i))))*(vard(i)/(1+vard(i)))
        part2(i) = ((vare(i)/(1+vare(i)))+(varf(i)/(1+varf(i)))+
     .                  (varg(i)/(1+varg(i))))*(varh(i)/(1+varh(i)))
        part3(i) = (vari(i)/(1+vari(i)))*(vard(i)/(1+vard(i)))
        stscor(i) = (part1(i)+part2(i)+part3(i))
10     continue
```

C) Finally, a small section of the new Fortran coding for the identification of inferior myocardial infarction is given.

```
c****  Set up the new smooth Q1 to QE4 ****

       do 660 i=1,7
         qq1(i) = ((pta5(i)*ptj1(i))+pta6(i)-(pta5(i)*ptj1(i)*pta6(i)))*
     .              (ptb3(i)*pti1(i))
         qq2(i) = ((pta5(i)*ptj1(i))+(pta4(i)*ptj3(i))-(pta5(i)*ptj1(i)*
     .                   pta4(i)*ptj3(i)))*(ptb5(i)*pti1(i))
         qq3(i) = (pta3(i)*ptj1(i)*ptb4(i)*pti1(i))
         qq4(i) = ((pta2(i)*ptj3(i))+(pta4(i)*pth2(i))-(pta2(i)*ptj3(i)*
     .               pta4(i)*pth2(i))*(pth1(i)+ptg1(i)-(pth1(i)*
     .                ptg1(i)))*(ptb2(i)*pti2(i))
         qqe1(i) = (ptc5(i)*ptk1(i))*(pti1(i)*(1-pte2(i)))*ptd2(i)*
     .            (ptg1(i)+pth2(i)-ptg1(i)*pth2(i))
         qqe2(i) = ((ptc5(i)*ptk1(i))+(ptc4(i)*ptk3(i))-(ptc5(i)*ptk1(i)*
     .              ptc4(i)*ptk3(i)))*(pti1(i)*(1-pte2(i)))*ptd4(i)*
     .             (ptg1(i)+pth2(i)-(ptg1(i)*pth2(i)))
         qqe3(i) = (ptc3(i)*ptk1(i))*pti1(i)*(1-pte2(i))*ptd3(i)*
     .            (ptg1(i)+pth2(i)-(ptg1(i)*pth2(i)))
         qqe4(i) = ((ptc2(i)*ptk3(i))+(ptc4(i)*pth2(i))-(ptc2(i)*ptk3(i)*
     .              ptc4(i)*pth2(i)))*pti2(i)*(1-pte2(i))*ptd1(i)*
     .             (ptg1(i)+pth2(i)-(ptg1(i)*pth2(i)))
660    continue
```