# Automatic Interpretation of

iew metadata, citation and similar papers at core.ac.uk

provid

**Peter R. Rijnbeek**

# Acknowledgements

Financial support by Welch Allyn Cardio Control B.V. for the publication of this thesis is gratefully acknowledged.

Rijnbeek, P.R.
Automatic Interpretation of Pediatric Electrocardiograms.
Thesis Erasmus University Rotterdam — with summary in Dutch.

ISBN: 978-90-9021557-0

This thesis was typeset by the author with LaTeX $2_\varepsilon$.
Cover design: J.P. Witteman

© P.R. Rijnbeek, 2007

# Automatic Interpretation of Pediatric Electrocardiograms

Automatische beoordeling van kinder-elektrocardiogrammen

**Proefschrift**

ter verkrijging van de graad van doctor
aan de Erasmus Universiteit Rotterdam
op gezag van de Rector Magnificus
Prof. dr. S.W.J. Lamberts
en volgens besluit van het College voor Promoties

de openbare verdediging zal plaatsvinden op
woensdag 7 maart 2007 om 15:45 uur

door

Peter Richard Rijnbeek
geboren te Voorburg

## Promotiecommissie

| | |
|---|---|
| Promotor: | Prof. dr. ir. J.H. van Bemmel |
| Overige leden: | Prof. dr. W.A. Helbing<br>Prof. dr. E.W. Steyerberg<br>Prof. dr. T. Stijnen |
| Co-promotor: | Dr. ir. J.A. Kors<br>Dr. M. Witsenburg |

Nederlandse Hartstichting
www.hartstichting.nl
Redt levens

Financial support by the Netherlands Heart Foundation for the publication of this thesis is gratefully acknowledged.

Voor mijn moeder

# Contents


**1 General Introduction** **1**

**2 Minimum bandwidth requirements for the recording of pediatric ECGs** **11**

**3 New normal limits for the pediatric electrocardiogram** **21**

**4 Exhaustive rule induction under constraints** **41**

**5 PEDMEANS: a computer program for the interpretation of pediatric ECGs** **69**

**6 Electrocardiographic criteria for left ventricular hypertrophy in children** **83**

**7 General Discussion** **97**

**Summary** **105**

**Samenvatting** **107**

**Curriculum Vitae** **111**

**Dankwoord** **113**

# 1 General Introduction

## 1.1 Computerized ECG interpretation

The year 1902 saw the birth of clinical electrocardiography when Willem Einthoven published the first electrocardiogram (ECG) of unprecedented quality recorded with his newly invented string- galvanometer [1]. The foundations of electrocardiographic diagnosis were laid in the half century that followed. After the second world war electronic pen-writing recorders made their appearance and quickly pushed the bulky string galvanometers from the scene, notwithstanding a far inferior frequency response. Standards for performance were then issued that were unfortunately based on the frequency characteristics of this type of equipment. We will return to this subject in the chapter on the minimum bandwidth requirements for the recording of pediatric ECGs (Chapter 2).

The first attempt to automate the analysis of the ECG by computer was made as early as in the late 1950s [2]. It was based on angles and magnitudes of spatial ventricular gradients, calculated from the time integrals over one PQRST complex in the three orthogonal leads. The X,Y and Z leads of the (Frank) vectorcardiogram (VCG) were used for the main reason that three leads lay less claim to computer capacity than 12 leads. Cycle recognition was obtained through analogue circuitry, and beginning and end of P, QRS and T were not identified in this simple screening program. The first automatic wave recognition program became available in 1961 [3]. This program, as well as its successors as developed by Pipberger, were also based on the vectorial XYZ leads. The first program for conventional 12-lead analysis was made by Caceres in 1962 [4]. By 1967 in the Netherlands Van Bemmel at the Institute of Medical Physics started the development of the Modular ECG Analysis System (MEANS), continued later in the departments of Medical Informatics of the Free University of Amsterdam and the Erasmus University of Rotterdam. In the present study all ECGs were analysed by MEANS, which has extensively been evaluated both by its developers [5] and by others [6].

Since these pioneering years and up to the present time a large research effort has been spent on improving the quality of computerized ECG analysis. Currently, ECG interpretation is offered in almost all electrocardiographs and PC-based ECG analysis systems.

The systems generally consist of a measurement part and a diagnostic interpretation part. The measurement part takes care of data acquisition, artifact detection and correction, followed by wave detection and determination of onsets and ends of the various waves. Finally, a set of measurements, such as wave amplitudes, durations, etc., are computated. The interpretation part derives the clinical significance of these mea-

surements. This results in a final classification, or interpretation, of the ECG in terms familiar to the physician.

Computerized ECG interpretation has a number of advantages. The computer analysis can improve the quality of the ECG recordings by, e.g., baseline correction, removal of mains interference, and artifact detection. Furthermore, it reduces inter- and intraobserver variability since it abolishes the subjective differences arising in visual interpretation and through a quantitative approach may enhance correct classification. Automatic storage and retrieval, with the possibility of comparing the new ECG with its predecessors (serial analysis) is another asset of computerized ECG processing.

However, the use of computer algorithms to interpret the ECG will only stand its ground in a clinical environment when the quality of the diagnostic interpretation has been proven. This instantiated a European effort, called 'Common Standards for Quantitative Electrocardiography' (CSE) [7]. This project was divided into two 5-year periods, the first dealing with measurements, the second with diagnostic interpretation. The main objectives of the first part of the CSE project were to establish standards for computer-derived ECG measurements and to compare the results of measurements from different programs. In the second part of the CSE project the diagnostic interpretations of various systems were compared with those of a panel of experienced cardiologists and with the diagnose based on clinical evidence, not involving the ECG. The CSE study showed that computer algorithms are almost as well as experienced electrocardiographers and can assist clinicians in interpreting ECGs. It made available reference libraries for the evaluation and improvement of ECG computer programs.

## 1.2 Pediatric ECG analysis

While much research has been devoted to computerized interpretation of the adult ECG, automatic interpretation of pediatric ECGs has received much less attention. Nevertheless, computer support in pediatric electrocardiography could be even more beneficial because the interpretation is more complex due to the strong age-dependency of the diagnostic criteria [8–10]. Especially in the first year of life the electrical behaviour of the heart changes rapidly [11]. This is mainly the result of a change from right to left ventricular predominance, which is reflected in changes in wave amplitudes. In addition, intervals are markedly reduced in the young child and, as the child grows, gradually prolong to adult values in relation to the decreasing heart rate. Consequently, for diagnostic criteria one has to rely on extensive tables of values for

amplitudes, durations, and angles. A computer program could be helpful in relieving the pediatric cardiologist from memorizing or consulting tables of large numbers of age-dependent criteria. In addition, qualified readers of pediatric ECGs are rare and mostly located in university centers [12]. Inter- and intraobserver variability in reading pediatric ECGs was shown to be substantial [13].

In the past the Mayo computer system routinely processed the pediatric VCG for some time [8], and in the VA Research Center for Cardiovascular Data, a large cooperative study has been undertaken in this field [14]. Other programs have been developed for pediatric VCG [15,16] and 12-lead ECG [17–19]. The combination of ECG and synthesized VCG measurements to discriminate between mild RVH with terminal conduction delay and partial RBBB in children was described by Zhou et al. [20]. Relatively few evaluation studies of pediatric ECG programs have been performed. In a study of 248 pediatric ECGs for which the clinical diagnosis was known by ECG-independent means, the HP pediatric program had a 70% sensitivity and 82% specificity for RVH and a 38% sensitivity and 93% specificity for LVH [21]. In a study by Hamilton et al. [22], the diagnoses of RVH and LVH by the pediatric version of the Glasgow program were compared with those of two pediatric cardiologists. When the cardiologists were not provided with clinical information, sensitivity of the program for RVH was 73% at a specificity of 97%, but sensitivity for LVH was only 25% at 96% specificity. Finally, pediatric ECG interpretations by the Marquette 12 SL program were compared with those of emergency department physicians in a 12-month prospective study, taking the interpretation by a pediatric electrophysiologist as the reference [23]. The computer proved to be more accurate than this group of physicians for interpretations considered to be of minimal or indeterminate clinical significance, but both performed poorly in interpreting the few cases that were classified as being of definite clinical significance by the expert.

## 1.3 Development of a pediatric ECG analysis system

We wanted to develop a pediatric version of MEANS, to be called the PEDiatric ECG Analysis System (PEDMEANS). For that purpose, we had to modify both the measurement part and the diagnostic interpretation part of MEANS.

The following issues are addressed in this thesis: minimum bandwidth requirements to record pediatric ECGs accurately, estimation of new up-to-date normal limits, formalization of expert knowledge with the use of a learning algorithm, and collection of a large pediatric ECG database for development and evaluation purposes.

### Minimum bandwidth requirements

For accurate recording of pediatric ECGs, the bandwidth of the acquisition system is of major importance. The bandwidth, defined as the frequency range between low and high-frequency cutoffs (-3 dB), should at least extend to the highest frequency component in the ECG signal. Previous studies that determined the frequency content had their limitations: the study population was too small or the sampling frequency used by the recording system was too low [24,25]. Therefore, we need to determine the minimum bandwidth requirements for pediatric ECGs based on a large set of ECGs recorded at a high sampling rate.

### Normal limits

For the development of a computer program for pediatric ECG interpretation, reliable normal limits are essential. Several studies have been conducted to this end [11,26–33]. However, all these studies have certain imperfections that limit their practical applicability. Firstly, normal limits have often been derived for an incomplete set of clinically relevant parameters and leads. Secondly, in many studies parameters were measured by hand from ECGs recorded on paper. Computer analysis of digitized ECGs allows more accurate measurement. Thirdly, in some studies the ECG signals may have been deformed owing to too low sampling rates or to the use of ECG amplifiers with too small bandwidth. Therefore, we want to establish up-to-date normal limits for the pediatric ECG based on a large set of normal ECGs analyzed by computer and recorded at a high sampling rate.

### Formalization of diagnostic criteria

The diagnostic interpretation part of PEDMEANS consists of diagnostic criteria represented in the form of decision rules for all pediatric ECG abnormalities. The elucidation of expert knowledge to obtain accurate decision rules is a tedious and time-consuming task. Whereas the pediatric cardiologists are generally well able to classify a ECG, they often have difficulties to articulate the knowledge they are using to do this, the more so if that knowledge has to be precisely formulated in order to be implemented in a computer program. To alleviate this problem, learning algorithms can be used that automatically construct a classifier based on a set of example cases. Each case consists of a set of feature values and a label indicating the class to which the case belongs. The learning algorithms search the feature space to generate the best classifier in terms of the number of correctly classified cases.

Most available algorithms aim to maximize the accuracy of the classifier that is being learned. However, in medicine, sensitivity and specificity are more commonly used measures to characterize classification performance. Often, a classifier needs to be maximal for one of these performance measures while fulfilling a user-specified minimum performance requirement for the other. Such optimization under constraints is difficult to achieve with available algorithms. Moreover, currently available algorithms cannot guarantee to find the best classifier by nature of their learning methodology. Therefore, research to develop an algorithm that overcomes these shortcomings is wanted.

### Data acquisition

No ECG reference databases, such as were established in the CSE project for adult ECGs [6], are available for the pediatric ECG. However, such a database is mandatory for program development and evaluation. A large number of pediatric ECGs has therefore to be collected from a hospital population with a diversity of abnormalities. These ECGs need to be interpreted by experienced pediatric cardiologists to obtain a reference diagnosis.

## 1.4 Aims and scope of this study

The aim of this study is to develop and evaluate a computer program for the interpretation of pediatric ECGs. In Chapter 2, we determine minimum bandwidth requirements for accurate recording of pediatric ECGs. The averaged beats of a large set of ECGs are passed through digital filters with different cutoff points and the absolute errors in maximum QRS amplitude for each simulated bandwidth are used to determine the minimum bandwidth requirements. In Chapter 3, we derive new normal limits for the pediatric ECG based on a large set of ECGS from children recruited at three child health centers, three primary schools, and a secondary school in Rotterdam. These new normal limits are used as cutoff points in the diagnostic criteria of the program. In Chapter 4, we describe research that extends our learning algorithm, called EXPLORE (Exhaustive Procedure for LOgic-Rule Extraction), that exhaustively generates rules that fulfill user-specified performance requirements [34]. Because of the exponential growth of the search space we present several new approaches that improve the speed of induction without loss of performance. In Chapter 5, we describe the modifications made in both the measurement part and diagnostic interpretation part of MEANS to allow pediatric ECG interpretation. We discuss the usage of

the learning algorithm in the development process and assess the performance of the system on an independent test set. Finally, in Chapter 6 we focus on the validity of the ECG in diagnosing left ventricular hypertrophy (LVH) in children and try to develop better criteria for computerized LVH diagnosis by combining diagnostic parameters with the use of the EXPLORE algorithm.

## 1.5 References

[1] Fye WB. A history of the origin, evolution, and impact of electrocardiography. Am J Cardiol 1994;73:937–49.

[2] Pipberger HV, Arms RJ, Stallmann FW. Automatic screening of normal and abnormal electrocardiograms by means of a digital electronic computer. Proc Soc Exp Biol Med 1961;106:130–132.

[3] Stallmann FW, Pipberge HV. Automatic recognition of electrocardiographic waves by digital computer. Circ Res 1961;9:1138–1143.

[4] Caceres CA, Steinberg CA, Abraham S, Carbery WJ, Mc BJ, Tolles WE, et al. Computer extraction of electrocardiographic parameters. Circulation 1962;25:356–62.

[5] Van Bemmel JH, Kors JA, Van Herpen G. Methodology of the modular ECG analysis system MEANS. Methods Inf Med 1990;29:346–53.

[6] Willems JL, Arnaud P, Van Bemmel JH, Bourdillon PJ, Degani R, Denis B, et al. A reference data base for multilead electrocardiographic computer measurement programs. J Am Coll Cardiol 1987;10:1313–21.

[7] Willems JL, Arnaud P, van Bemmel JH, Degani R, Macfarlane PW, Zywietz C. Common standards for quantitative electrocardiography: goals and main results. CSE Working Party. Methods Inf Med 1990;29:263–71.

[8] Guller B, O'Brien PC, Smith RE, Weidman WH, DuShane JW. Computer interpretation of Frank vectorcardiogram in normal infant: Longitudinal and cross-sectional observations from birth to 2 years of age. J Electrocardiol 1975;8:201–8.

[9] Hamilton KH, Clifton JF, Laks MM. Computerized pediatric ECG interpretation. A review of the first year of live. In: Pryor A, Bailey J, editors. Computerized interpretation of the ECG IV. New York: Engineering Foundation, 1979; pp. 239–312.

[10] Brohet CR, Derwael-Barchy C, Robert A, Fesler R, Styns M, Brasseur LA, et al. Computer interpretation of pediatric Frank vectorcardiograms in the evaluation of congenital heart disease. Am J Cardiol 1983;52:127–32.

[11] Davignon A, Rautaharju P, Boisselle E, Soumis F, Megelas M, Choguette A. Normal ECG standards for infants and children. Pediatr Cardiol 1979/80;1:123–31.

[12] Horton LA, Mosee S, Brenner J. Use of the electrocardiogram in a pediatric emergency department. Arch Pediatr Adolesc Med 1994;148:184–188.

[13] Hamilton RM, McLeod K, Houston AB, Macfarlane PW. Inter- and intraobserver variability in LVH and RVH reporting in pediatric ECGs. Ann Noninvasive Electrocardiol 2005; 10:330–333.

[14] Guller B, Lau FY, Dunn RA, Pipberger HA, Pipberger HV. Computer analysis of changes in Frank vectorcardiograms of 666 normal infants in the first 72 hours of life. J Electrocardiol 1977;10:19–26.

[15] Zywietz C, Borovsky D, Gotsch G, Joseph G. Methodology of ECG interpretation in the Hannover program. Methods Inf Med 1990;29:375–385.

[16] Brohet CR, Robert A, Derwael C, Fesler R, Stijns M, Vliers A. Computer interpretation of pediatric orthogonal electrocardiograms: statistical and deterministic classification methods. Circulation 1984;70:255–262.

[17] Francis DB, Miller BL, Benson D. W. J. A new computer program for the analysis of pediatric scalar electrocardiograms. Comput Biomed Res 1981;14:63–77.

[18] Laks MM. A computer program for intepretation of infant and children electrocardiograms. In: Willems JL, van Bemmel JH, Zywietz C, editors. Computer ECG Analysis: Towards Standardization. Amsterdam: North-Holland, 1986; pp. 59–65.

[19] Macfarlane PW, Coleman EN, Devine B, Houston A, McLaughlin S, Aitchison TC, et al. A new 12-lead pediatric ECG interpretation program. J Electrocardiol 1990;23 Suppl:76–81.

[20] Zhou SH, Liebman J, Dubin AM, Gillette PC, Gregg RE, Helfenbein ED, et al. Using 12-lead ECG and synthesized VCG in detection of right ventricular hypertrophy with terminal right conduction delay versus partial right bundle branch block in the pediatric population. J Electrocardiol 2001;34 Suppl:249–57.

[21] Guller B, Jones T, McCloskey J, Herndon SP. The Hewlett-Packard pediatric ECG computer program (HP-P3) and independent clinical information. J Electrocardiol 1990;23 Suppl:204.

[22] Hamilton RM, Houston AB, McLeod K, Macfarlane PW. Evaluation of pediatric electrocardiogram diagnosis of ventricular hypertrophy by computer program compared with cardiologists. Pediatr Cardiol 2005;26:373–8.

[23] Snyder CS, Fenrich AL, Friedman RA, Macias C, O'Reilly K, Kertesz NJ. The emergency department versus the computer: which is the better electrocardiographer? Pediatr Cardiol 2003;24:364–8.

[24] Riggs T, Isenstein B, Thomas C. Spectral analysis of the normal electrocardiogram in children and adults. J Electrocardiol 1979;12:377–9.

[25] Berson AS, Lau FY, Wojick JM, Pipberger HV. Distortions in infant electrocardiograms caused by inadequate high- frequency response. Am Heart J 1977;93:730–4.

[26] Maroney M, Rantz LA. Electrocardiograms in 679 healthy infants and children. Pediatrics 1950;5:396.

[27] Ziegler RF. Electrocardiographic studies in normal infants and children. Springfield: Charles C. Thomas, 1951.

[28] Strong WB, Downs TD, Liebman J, Liebowitz R. The normal adolescent electrocardiogram. Am Heart J 1972;83:115–28.

[29] Liebman J, Plonsey R, Gillette PC. Pediatric cardiology. Baltimore: Williams & Wilkins, 1982.

[30] Macfarlane PW, Coleman EN, Pomphrey EO, McLaughlin SC, Houston A, Aitchison T. Normal limits of the high-fidelity pediatric ECG. Preliminary observations. J Electrocardiol 1989;22:162–8.

[31] Macfarlane PW, McLaughlin SC, Devine B, Yang TF. Effects of age, sex, and race on ECG interval measurements. J Electrocardiol 1994;27:14–9.

[32] Pearl W. Effects of gender, age, and heart rate on QT intervals in children. Pediatr Cardiol 1996;17:135–6.

[33] Tutar HE, Ocal B, Imamoglu A, Atalay S. Dispersion of QT and QTc interval in healthy children, and effects of sinus arrhythmia on QT dispersion. Heart 1998;80:77–9.

[34] Kors JA, Hoffmann AL. Induction of decision rules that fulfil user-specified performance requirements. Pattern Recognition Letters 1997;19:1187–1195.

# 2 Minimum bandwidth requirements for the recording of pediatric ECGs

P.R. Rijnbeek*, M. Witsenburg†, A. Szatmari‡, J. Hess§, J.A. Kors*

* Dept. of Medical Informatics, Faculty of Medicine and Health Sciences, Erasmus University, Rotterdam, The Netherlands
† Dept. of Pediatric Cardiology, Sophia Children's Hospital, Rotterdam, The Netherlands
‡ Dept. of Pediatric Cardiology, Hungarian Institute of Cardiology, Budapest, Hungary
§ Dept. of Pediatric Cardiology and Congenital Heart Disease, Deutsches Herzzentrum München, Munich, Germany

Circulation 2001;104:3087-3090

## Abstract

Previous studies that determined the frequency content of the pediatric electrocardiogram (ECG) had their limitations: the study population was small or the sampling frequency used by the recording system was low. Therefore, current bandwidth recommendations for recording pediatric ECGs are not well founded. We wanted to establish minimum bandwidth requirements with the use of a large set of pediatric ECGs recorded at a high sampling rate.

For 2169 children aged 1 day to 16 years, a 12-lead ECG was recorded at a sampling rate of 1200 Hz. The averaged beats of each ECG were passed through digital filters with different cutoff points (50-300 Hz in 25 Hz steps). We measured the absolute errors in maximum QRS amplitude for each simulated bandwidth, and determined the percentage of records with an error greater than 25 μV. We found that in any lead a bandwidth of 250 Hz yields amplitude errors less than 25 μV in more than 95% of the children younger than one year. For older children, a gradual decrease in ECG frequency content was demonstrated.

We recommend a minimum bandwidth of 250 Hz for the recording of pediatric ECGs. This bandwidth is considerably higher than the previous recommendation of 150 Hz of the American Heart Association.


Keywords: electrocardiography, pediatrics, bandwidth

## 2.1 Introduction

For accurate recording of electrocardiograms (ECGs), the bandwidth of the recording system is of major importance. The bandwidth, defined as the frequency range between low and high-frequency cutoffs (-3 dB), should at least extend to the highest frequency component in the ECG signal. Many studies have been performed to determine the frequency content of the adult ECG [1–8], but only few studies addressed the frequency content of the pediatric ECG [9,10].

Riggs et al. [9] studied the frequency spectrum of the ECG of only 8 children and concluded that the vast majority of the information is confined to frequencies below 100 Hz. However, determining the frequency content of an ECG by inspecting its frequency spectrum is difficult. The ECG spectra appear to be monotonically decreasing, and it is hard to tell which frequency components belong to the ECG and which to noise. Moreover, this approach does not show the effect of reduced bandwidth on wave amplitudes, which from a clinical point of view is the more important information.

Berson et al. [10] recorded vectorcardiograms from a group of 600 infants. They filtered the waveforms with different low-pass filters and determined the amplitude differences between filtered and unfiltered waves, concluding that a bandwidth of 100 Hz is required to avoid amplitude errors of 10% or more. However, the original signals were recorded at a sampling rate of 500 Hz. It has been questioned whether this rate is high enough to obtain accurate measurements in pediatric ECGs [1,4,11].

The American Heart Association (AHA) recommends 150 Hz as minimum bandwidth and 500 Hz as minimum sampling rate for the recording of pediatric ECGs, but also states that it is yet unknown how far the bandwidth of systems has to extend, due to the limitations of previous studies [12]. In the present study, we wanted to determine the minimum bandwidth requirements for recording pediatric ECGs with the use of a large set of ECGs recorded at a high sampling rate.

## 2.2 Methods

The study population consisted of 2169 children, with and without cardiac abnormalities, who had been referred to the pediatric cardiology department of the Sophia Children's Hospital in Rotterdam, the Netherlands. Table 2.1 shows the age distribution of the children. For each child, a 12-lead ECG was recorded at a sampling rate

of 1200 Hz using a PC-based acquisition system (Cardio Control, Delft, the Netherlands). The frequency response of the ECG recorder was flat to 320 Hz (-3 dB point). Following common practice in the department of pediatric cardiology in Rotterdam, $V_{3R}$ was used instead of $V_3$ and $V_7$ instead of $V_5$.

**Table 2.1.** Age distribution of the children.

| Age | Total |
|---|---|
| 0 to 3 month | 606 |
| 3 to 6 months | 226 |
| 6 to 12 months | 213 |
| 1 to 3 years | 244 |
| 3 to 5 years | 240 |
| 5 to 8 years | 265 |
| 8 to 12 years | 189 |
| 12 to 16 years | 186 |
| Total | 2169 |

All ECGs were processed by the Modular Ecg ANalysis System (MEANS), which has extensively been evaluated both by its developers [13] and by others [14]. For each of the twelve leads, the program computes a representative averaged beat, from which ECG measurements are derived. The averaged beats of the 1200 Hz recordings were passed through digital low-pass filters with different -3 dB points, at 50, 75, to 300 Hz in 25 Hz steps, thus simulating recording systems with reduced bandwidths. The filters were developed with the signal processing toolbox of Matlab (Massachusetts, USA). They were designed to have a ripple smaller than 0.001 dB in the pass-band, and an attenuation of less than -40 dB within 25 Hz of the cutoff point.

To determine the effect of reduced bandwidth on wave amplitudes, we measured the absolute differences between the maximum QRS deflections in the filtered and unfiltered leads for each of the cutoff points, and calculated the 95th percentile of these absolute differences. Since we expected the frequency content of the pediatric ECG to decrease with increasing age, ECGs from children younger than one year (n = 1045) were taken to determine minimum bandwidth requirements. The total population was used to assess the effect of age on the frequency content of the ECG signals. Additionally, we calculated the percentage of recordings in which the absolute differences between the maximum deflections of the filtered and unfiltered leads exceeded 25 μV. The 25 μV threshold was chosen because this was considered an amplitude difference still distinguishable by human interpreters from standard paper ECG recordings.

## 2.3 Results

Fig. 2.1 and Fig. 2.2 show the 95th percentile of the absolute differences between filtered and unfiltered leads for the maximum positive and maximum negative QRS deflections, respectively.

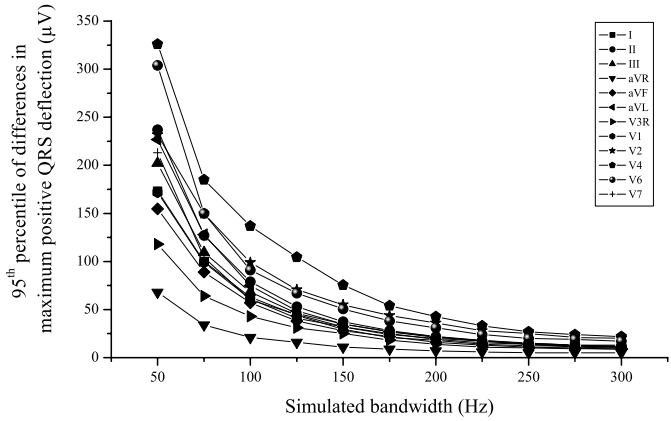


**Figure 2.1.** 95th percentile of the absolute differences between the maximum positive QRS deflection in filtered and unfiltered leads for different bandwidths.

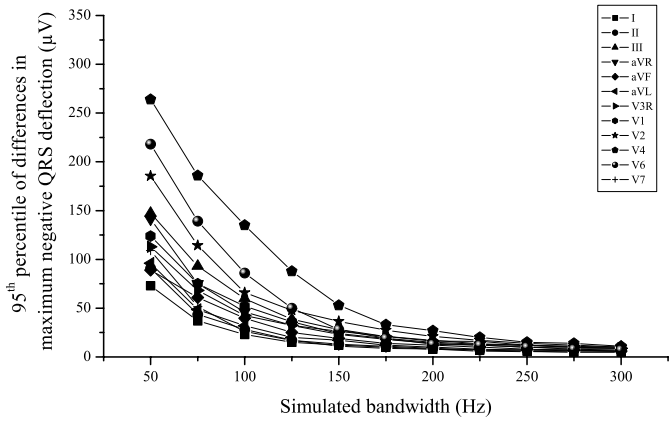


**Figure 2.2.** 95th percentile of the absolute differences between the maximum negative QRS deflection in filtered and unfiltered leads for different bandwidths.

The largest absolute differences in both positive and negative deflections are found in lead $V_4$, whereas the smallest differences are seen in lead aVR for positive deflections and in lead I for negative deflections. The clinically important leads $V_2$ and $V_6$ also show large absolute differences in both positive and negative deflections.

Fig. 2.3 presents, for each lead, the percentage of recordings with the differences between the maximum positive QRS deflections in the filtered and unfiltered leads exceeding 25 μV. Fig. 2.4 gives these results for maximum negative deflections. The errors in the positive deflections are higher than in the negative deflections and amplitude errors increase considerably with decreasing bandwidth. In any lead, a bandwidth of 250 Hz yields amplitude errors less than 25 $\mu V$ in more than 95% of the cases (see Fig. 2.1 and Fig. 2.3).

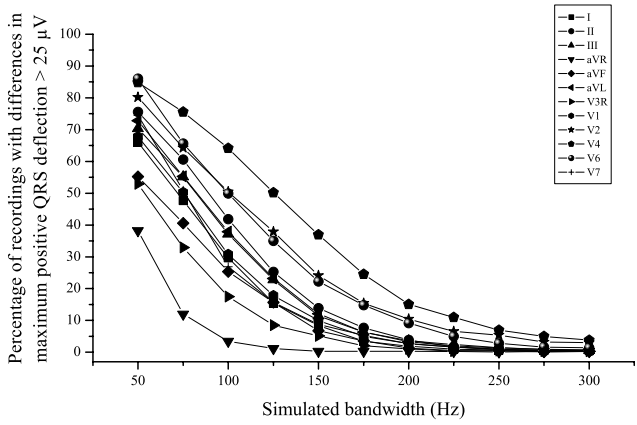


**Figure 2.3.** Percentage of recordings with the differences of the maximum positive QRS deflection in filtered and unfiltered leads exceeding 25 μV for different bandwidths.

Fig. 2.5 shows the relationship between age and the 95th percentile of the difference in maximum positive QRS deflections in lead $V_1$ for different bandwidths, illustrating that ECGs have a higher frequency content at younger ages. For example, for children from 0 to 3 months the effect of a low-pass filter at 100 Hz triples as compared to the oldest children in our study population. In $V_4$, the lead that showed the largest amplitude differences, there is a small initial increase of the difference with increasing age up to approximately one year, with a gradual decrease onwards (Fig. 2.6). The other leads show comparable patterns of decreasing frequency content with increasing age.

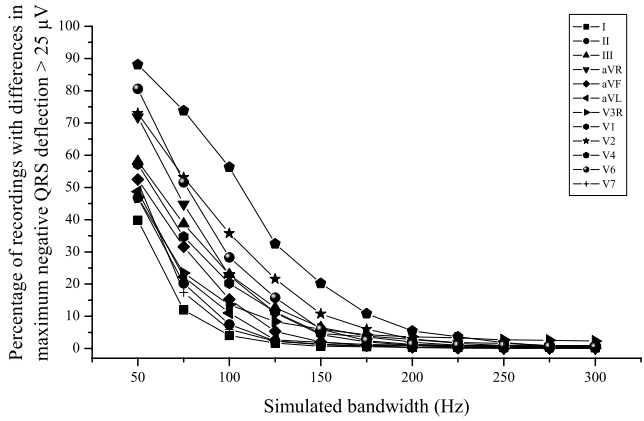


**Figure 2.4.** Percentage of recordings with the differences of the maximum negative QRS deflection in filtered and unfiltered leads exceeding 25 µV for different bandwidths.

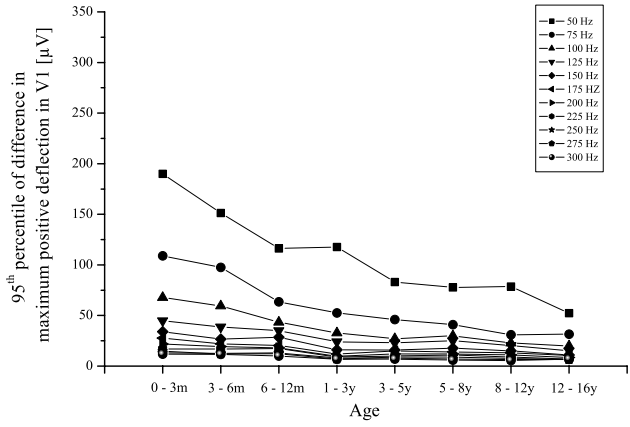


**Figure 2.5.** 95th percentile of the differences of the maximum positive QRS deflection in the filtered and unfiltered lead $V_1$ for different age groups.

## 2.4 Discussion

To establish minimum bandwidth requirements for accurate recording of pediatric ECGs, we used a large set of pediatric ECGs recorded at a high sampling frequency of 1200 Hz, thus obviating some of the limitations of previous studies [9,10]. To our

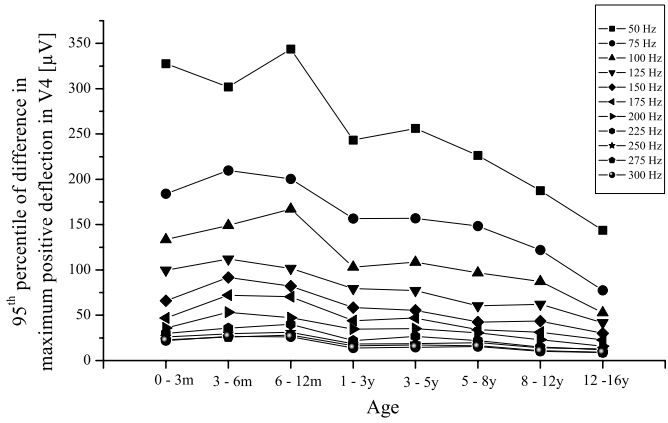


**Figure 2.6.** 95th percentile of the differences of the maximum positive QRS deflection in the filtered and unfiltered lead $V_4$ for different age groups.

knowledge, this is the first study that demonstrates the effect of bandwidth limitations for all 12 leads of the ECG and that illustrates the influence of age on frequency content. Based on our results, we recommend a minimum bandwidth of 250 Hz for the recording of pediatric ECGs. This bandwidth requirement of 250 Hz is considerably higher than the 150 Hz previously recommended by the AHA [12]. With a bandwidth of 150 Hz, 38% of the cases in our study has an error greater than 25 μV in the maximum positive deflection in lead $V_4$. For leads $V_2$ and $V_6$, these percentages are 25% and 23%, respectively. In vectorcardiographic leads, Berson et al. [10] found amplitude errors greater than 50 μV in 8% of the R-wave amplitudes and 5% of the S-wave amplitudes when using a 150 Hz filter. In our study, 15% of the positive deflections and 7% of the negative deflections in $V_4$ have amplitude errors exceeding 50 μV when a 150 Hz filter is used. These differences between the two studies may partly be explained by the difference in sampling rate (500 vs 1200 Hz) and by the use of different lead systems. Furthermore, the analyses by Berson et al. were not done on each separate lead but on leads X, Y, and Z combined, which is likely to underestimate the effect of filtering for individual leads. More importantly, however, we find a threshold of 25 μV instead of 50 μV is to be prefered for measuring the effect of reduced bandwidth on signal amplitudes.

We also studied the effect of age on the frequency content of the ECG signals. As shown in Fig. 2.5 and Fig. 2.6, the frequency content gradually decreases from infancy till adulthood. Strictly speaking for older children a lower bandwidth would suffice.

Our data for children of 12 to 16 years indicate that the system bandwidth should be about 150 Hz to yield amplitude errors less than 25 μV in more than 95% of the cases in this age group. This is close to the 125 Hz recommendation of the AHA for the adult ECG [12]. Nevertheless, we recommend to use the minimum bandwidth of 250 Hz for the entire pediatric population, because the age-range of the patients will often not be known in advance.

An increase in bandwidth will also affect the sampling rate. Shannon's theorem prescribes a minimum sampling rate that is twice the highest frequency in a signal, which in our case gives a rate of 500 Hz. However, this theorem is valid only for an infinite sampling period and would require a sophisticated but impractical interpolation technique. Therefore, the AHA recommends a sampling rates of 2 or 3 times the theoretical minimum [12]. On the basis of this rule of thumb, a sampling frequency of at least 1000 Hz would seem desirable.

## 2.5 References

[1] Barr RC, Spach MS. Sampling rates required for digital recording of intracellular and extracellular cardiac potentials. Circulation 1977;55:40–8.

[2] Berson AS, Pipberger HV. Electrocardiographic distortions caused by inadequate high-frequency response of direct-writing electrocardiographs. Am Heart J 1967;74:208–18.

[3] Berson AS, Ferguson TA, Batchlor CD, Dunn RA, Pipberger HV. Filtering and sampling for electrocardiographic data processing. Comput Biomed Res 1977;10:605–16.

[4] Golden DP, Wolthuis RA, Hoffler GW. A spectral analysis of the normal resting electrocardiogram. IEEE Trans Biomed Eng 1973;20:366–72.

[5] Pizzuti GP, Cifaldi S, Nolfe G. Digital sampling rate and ECG analysis. J Biomed Eng 1985; 7:247–50.

[6] Scher AM, Young A. Frequency analysis of electrocardiograms. Circulation Research 1960; 8:344–6.

[7] Yamamoto H, Miyahara H, Domae A. Is a higher sampling rate desirable in the computer processing of the pediatric electrocardiogram? J Electrocardiol 1987;20:321–8.

[8] Seegobin RD, Mohamed SA. Sex differences and frequency content of the Frank lead signal-averaged ECG in a normal population. J Electrocardiol 1996;29:105–9.

[9] Riggs T, Isenstein B, Thomas C. Spectral analysis of the normal electrocardiogram in children and adults. J Electrocardiol 1979;12:377–9.

[10] Berson AS, Lau FY, Wojick JM, Pipberger HV. Distortions in infant electrocardiograms caused by inadequate high- frequency response. Am Heart J 1977;93:730–4.

[11] Garson A. Clinically significant differences between the "old" analog and the "new" digital electrocardiograms. Am Heart J 1987;114:194–7.

[12] Bailey JJ, Berson AS, Garson A, Horan LG, Macfarlane PW, Mortara DW, et al. Recommendations for standardization and specifications in automated electrocardiography: bandwidth and digital signal processing. Circulation 1990;81:730–9.

[13] Van Bemmel JH, Kors JA, Van Herpen G. Methodology of the modular ECG analysis system MEANS. Methods Inf Med 1990;29:346–53.

[14] Willems JL, Arnaud P, Van Bemmel JH, Bourdillon PJ, Degani R, Denis B, et al. A reference data base for multilead electrocardiographic computer measurement programs. J Am Coll Cardiol 1987;10:1313–21.

# 3 New normal limits for the pediatric electrocardiogram

P.R. Rijnbeek*, M. Witsenburg†, A. Szatmari‡, J. Hess§, J.A. Kors*

* Dept. of Medical Informatics, Faculty of Medicine and Health Sciences, Erasmus University, Rotterdam, The Netherlands
† Dept. of Pediatric Cardiology, Sophia Children's Hospital, Rotterdam, The Netherlands
‡ Dept. of Pediatric Cardiology, Hungarian Institute of Cardiology, Budapest, Hungary
§ Dept. of Pediatric Cardiology and Congenital Heart Disease, Deutsches Herzzentrum München, Munich, Germany

European Heart Journal 2001;22:702-711

## Abstract

Previous studies that determined normal limits for the pediatric electrocardiogram (ECG) had their imperfections: ECGs were recorded with relatively low sampling rate, ECG measurements were done manually, or normal limits were presented for only a limited set of parameters. The aim of this study was to establish an up-to-date and complete set of clinically relevant normal limits for the pediatric ECG.

ECGs from 1912 healthy Dutch children (age 11 days to 16 years) were recorded at a sampling rate of 1200 Hz. The digitally stored ECGs were analysed using a well-validated ECG computer program. Normal limits of all clinically relevant ECG measurements were determined for nine age groups. Clinically significant differences were shown to exist, compared with previously established normal limits. Sex differences could be demonstrated for QRS duration and several amplitude measurements.

These new normal limits differ substantially from those commonly used and suggest that diagnostic criteria for the pediatric ECG should be adjusted.


Keywords: electrocardiography, pediatrics, normal limits, computer

## 3.1 Introduction

Several studies have been conducted to determine normal limits for the pediatric electrocardiogram (ECG) [1–9]. However, all these studies have certain imperfections that limit their practical applicability. Firstly, normal limits have often been presented for an incomplete set of clinically relevant parameters and leads. Secondly, in many studies parameters were measured by hand from ECGs recorded on paper. At present, computer analysis of digitised ECGs allows more accurate measurement. Thirdly, in some studies the ECG signals may have been recorded less than perfectly owing to low sampling rates or the use of ECG amplifiers with small bandwidth.

Probably the most comprehensive study to date has been that of Davignon et al. [4], in which ECGs of 2141 children aged 0 to 16 years were recorded. The ECGs were digitised at a sampling rate of 333 Hz and normal limits were determined using computer-assisted measurements. Normal limits for a large number of parameters were presented as percentile charts ranging from the 2nd to the 98th percentile. However, in a study of 1780 children, Macfarlane et al. [6] recorded ECGs at a sampling rate of 500 Hz and showed that the 98th percentile of normal amplitudes could be up to 46% higher than published by Davignon [4]. Unfortunately, in the study of Macfarlane normal limits were presented for only a few parameters [6] . Moreover, it has been questioned whether even a sampling rate of 500 Hz is high enough to obtain accurate measurements in pediatric ECGs [10–12].

In this study, we wanted to establish an up-to-date and complete set of clinically relevant normal limits for the pediatric ECG, using a high sampling rate of 1200 Hz and an ECG computer program for measurement.

## 3.2 Methods

### Study population

The study population consisted of 1912 children aged 11 days to 16 years, recruited at three child health centers, three primary schools, and a secondary school in Rotterdam. In the Netherlands all children as from about three weeks to four years old, periodically visit child health centers for a physical examination. Children with previously known cardiovascular abnormalities were excluded from the study. The total population is divided into nine age groups, similar to the grouping used by Davignon et al [4]. All children up to one month are combined in one group, because of the relatively small sample size. Table 3.1 shows the sex distribution for each age group.

For each child, weight and height was measured prior to the ECG recordings. Data for weight and height corresponded well with the Dutch growth standard [13]. The study was approved by the local ethics committee. All parents and all children aged 12 or older gave their written informed consent.

**Table 3.1.** Age and sex distribution of the study population.

| Age * | Male | Female | Total |
|---|---|---|---|
| 0 to 1 month | 16 | 28 | 44 |
| 1 to 3 months | 67 | 71 | 138 |
| 3 to 6 months | 78 | 104 | 182 |
| 6 to 12 months | 130 | 105 | 235 |
| 1 to 3 years | 95 | 110 | 205 |
| 3 to 5 years | 79 | 79 | 158 |
| 5 to 8 years | 142 | 118 | 260 |
| 8 to 12 years | 137 | 187 | 324 |
| 12 to 16 years | 200 | 166 | 366 |
| Total | 944 | 968 | 1912 |

* The term "to" specifies the upper limit of the age range in the sense of "less than" logic.

**ECG measurements**

For each child, a 12-lead ECG was recorded using a portable PC-based acquisition system (Cardio Control, Rijswijk, the Netherlands) at a sampling rate of 1200 Hz. The frequency response of this recorder is flat to 320 Hz (-3 db point). The ECGs were recorded by the same technician throughout the study. Following common practice in the department of pediatric cardiology in Rotterdam, $V_{3R}$ was used instead of $V_3$ and $V_7$ instead of $V_5$. All ECGs were processed by the Modular Ecg ANalysis System (MEANS) [14]. To reduce noise, MEANS computes a representative average beat for each of the twelve leads, from which ECG measurements are derived. MEANS has extensively been evaluated both by its developers [14] and by others [15,16]. In the latter studies, the performance of MEANS was gauged against the measurements obtained from a group of cardiologists, and its good performance shown. Plots of all

ECGs showing wave onsets and ends as found by MEANS were visually checked. Because of waveform recognition errors, mainly due to excessive noise, 16 ECGs were removed from the data set. The excluded ECGs were randomly distributed over the age groups.

**Estimation of normal limits**

The 2nd and 98th percentile of the measurement distribution were taken as the lower limit and the upper limit of normal, respectively. Zero amplitude values indicating absent Q, R, or S waves, were excluded from the statistical analysis of the data. Prior to the estimation of the percentiles, a linear regression analysis on age was performed in each age group. Percentiles were then estimated parametrically on the residuals. Since parametric estimation assumes a sample distribution to be gaussian, possible non-gaussianity of the residuals was removed using a two-stage transformation procedure as recommended by Solberg [17]. In the first stage, asymmetry (skewness) was iteratively eliminated using the exponential function of Manly [18]. In the second stage, peakedness (kurtosis) of the resulting symmetrical distribution was iteratively eliminated with the modulus function of John [19]. To test the gaussianity of the transformed distribution, the Kolmogorov-Smirnov test was used. Finally, estimated percentiles and their 95% confidence intervals were back-transformed to the original unit of measurement. If a distribution remained non-gaussian after transformation, the non-parametrical ranked-based method as described by Solberg [17] was applied on the original data. Sex differences were identified by non-overlapping 95% confidence intervals of the percentiles.

Apart from a tabular presentation of normal limits in age groups, we determined age-dependent curves that present the normal limits in a continuous form. The two-stage transformation as described above was applied in a window of 200 measurements, moving along the age axis with a step size of one measurement. For each window position the percentiles and their confidence intervals were calculated and related to the median of the age values included in the window. This procedure would imply that the first normal limit corresponds with the median age of the first 200 measurements. To allow for estimates at younger ages, the procedure starts with a small initial window that grows till 200 measurements are included. As a consequence, confidence intervals at the youngest ages are wider. Polynomial curves were then fitted through the 2nd and 98th percentile values to obtain percentiles that smoothly change with age. The order of the polynomials was determined by visual inspection of the fit, selecting the lowest order that yielded curves remaining within the estimated confidence intervals.

## 3.3 Results

Table 3.2 to Table 3.7 on the following pages show normal limits for the clinically most relevant parameters. Median values are given together with the 98th percentiles, taken as the upper limits of normal (ULN). The 2nd percentiles, taken as the lower limits of normal (LLN), are supplemented if clinically relevant. Normal limits are presented separately for boys (upper row) and girls (lower row). To indicate sex differences, non-overlapping 95% confidence intervals are visualized by bold percentiles. Because of space limitations, continuous age-dependent percentile curves are only shown for heart rate (Fig. 3.1) and QRS duration (Fig. 3.2) on page 33.

Table 3.2 summarizes the normal limits for the lead-independent ECG measurements. Heart rate substantially decreases with age as also illustrated in Fig. 3.1. The ULN of the heart rate is slightly higher for girls than for boys from the age of 8 years onward. The decrease in heart rate during growth is accompanied by an increase in the duration of the P wave, PR interval, and QRS complex. The median QRS duration is greater for boys than for girls in most age groups, but the differences in ULNs are small, ranging from 2 to 7 ms. The median QRS axis is directed to the right in the first months of life, reflecting the still increased right ventricular mass in that period. From the age of 3-6 months no further changes in QRS-axis direction are observed. The QTc interval, calculated according to Bazett's formula [20], remains relatively constant over the years with an ULN of approximately 450 ms.

In Table 3.3, the P-wave amplitude is given for leads II, $V_1$, and $V_2$. The P-wave amplitudes in II and $V_1$ do not change during growth, while in $V_2$ a gradual decrease with age is apparent. The highest ULNs of the P-wave amplitude, approximately 0.25 mV, were found in lead II.

The Q-wave amplitude is presented for clinically important leads in Table 3.4. The ULN of the Q-wave amplitude in the first month of life increases at least twofold to a maximum between one and three years, after which a decrease is seen towards the initial value. In the 12-16 year group, girls have significantly lower ULNs of the Q-wave amplitude in $V_6$ and $V_7$ than boys.

The normal limits of the amplitude of the R and S wave are shown in Table 3.5 and 3.6, respectively. R-wave amplitudes decrease with age in the right precordial leads, with a concomitant increase in the left precordial leads. S-wave amplitudes show a similar but inverse pattern. In the early adolescent years, girls have substantially lower precordial R-wave amplitudes than boys. However, the S waves in $V_4$, $V_6$, and $V_7$ are lower for girls than for boys from the first month of age onward.

**Table 3.2.** Lead-independent ECG measurements for boys (upper row) and girls (lower row): median (2nd percentile, 98th percentile)

| Lead | 0-1 months | 1-3 months | 3-6 months | 6-12 months | 1-3 years | 3-5 years | 5-8 years | 8-12 years | 12-16 years |
|---|---|---|---|---|---|---|---|---|---|
| Heart rate (bpm) | 160 (129,192) | 152 (126, 187) | 134 (**112,165**)* | 128 (106,194) | **119** (97,**155**) | 98 (73,123) | 88 (62,113) | 78 (55,**101**) | 73 (**48,99**) |
| | 155 (136,216) | 154 (126,200) | 139 (**122,191**) | 134 (106,187) | **128** (95,**178**) | 101 (78,124) | 89 (68,115) | 80 (58,**110**) | 76 (**54,107**) |
| P axis (°) | 56 (13,99) | 52 (10,73) | 49 ( -5,70) | 49 ( 9,87) | 48 (-12,78) | 43 (-13,69) | 41 (-54,72) | 39 (-17,76) | 40 (-24,76) |
| | 52 (24,80) | 48 (20,77) | 51 (16,80) | 50 (14,69) | 47 ( 1,90) | 44 ( -6,90) | 42 (-13,77) | 42 (-15,82) | 45 (-18,77) |
| P duration (ms) | 78 (64,85) | 79 (65,98) | 81 (64,103) | 80 (66,96) | 80 (63,113) | 87 (67,102) | 92 (73,108) | 98 (78,117) | 100 (82,118) |
| | 79 (69,106) | 78 (62,105) | 78 (63,106) | 80 (64,07) | 83 (62,104) | 84 (66,101) | 89 (71,107) | 94 (75,114) | 98 (78,122) |
| PR interval (ms) | 99 (77,120) | 98 (85,120) | 106 (87,134) | 114 (82,141) | 118 (86,151) | 121 (98,152) | 129 (99,160) | 134 (105,174) | 139 (107,178) |
| | 101 (91,121) | 99 (78,133) | 106 (84,127) | 109 (88,133) | 113 (78,147) | 123 (99,153) | 124 (92,156) | 129 (103,163) | 135 (106,176) |
| QRS axis (° ) | 97 (75,140) | 87 (37,138) | 66 ( -6,107) | 68 (14,122) | 64 (-4,118) | 70 (7,112) | 70 (-10,112) | 70 (-21,114) | 65 (-9,112) |
| | 110 (63,155) | 80 (39,121) | 70 (17,108) | 67 ( 1,102) | 69 ( 2,121) | 69 (3,106) | 74 ( 27,117) | 66 ( 5,117) | 66 ( 5,101) |
| QRS duration (ms) | 67 (50,85) | 64 (52,77) | 66 (54,**85**) | **69** (52,**86**) | **71** (54,88) | **75** (58,92) | **80** (63,98) | **85** (67,103) | **91** (**78**,111) |
| | 67 (54,79) | 63 (48,77) | 64 (50,**78**) | **64** (52,**80**) | **68** (54,85) | **71** (58,88) | **77** (59,95) | **82** (66,99) | **87** (**72**,106) |
| QTc interval (ms)† | 413 (378,448) | 419 (**396**,458) | 422 (391,453) | 411 (379,449) | 412 (383,455) | 412 (377,448) | 411 (371,443) | 411 (373,440) | 407 (362,449) |
| | 420 (379,462) | 424 (**381**,454) | 418 (386,448) | 414 (381,446) | 417 (381,447) | 415 (388,442) | 409 (375,449) | 410 (365,447) | 414 (370,457) |

* Bold values indicate that the 95% confidence intervals of the percentile estimates for boys and girls do not overlap.

† Corrected QT interval, according to Bazett's formula: $QT_c = QT \times \sqrt{heart\ rate/60}$.

**Table 3.3.** P-wave amplitudes (mV) for boys (upper row) and girls (lower row): median (98th percentile)

| Lead | 0-1 m | 1-3 m | 3-6 m | 6-12 m | 1-3 y | 3-5 y | 5-8 y | 8-12 y | 12-16 y |
|---|---|---|---|---|---|---|---|---|---|
| II | 0.15 (0.23) | 0.16 (0.25) | 0.15 (0.22) | 0.16 (0.26) | 0.15 (0.25) | 0.13 (0.23) | 0.12 (0.22) | **0.12** (0.22)* | **0.13** (0.24) |
| | 0.16 (0.25) | 0.16 (0.23) | 0.16 (0.23) | 0.16 (0.24) | 0.16 (0.25) | 0.13 (0.23) | 0.12 (0.24) | **0.14** (0.24) | **0.15** (0.26) |
| $V_1$ | 0.12 (0.22) | 0.10 (0.19) | 0.09 (**0.17**) | 0.10 (0.17) | 0.13 (0.20) | 0.12 (0.19) | 0.12 (0.20) | 0.11 (0.19) | 0.11 (0.18) |
| | 0.10 (0.19) | 0.10 (0.16) | 0.10 (0.16) | 0.11 (**0.21**) | 0.12 (0.20) | 0.12 (0.20) | 0.11 (0.18) | 0.11 (0.19) | 0.10 (0.17) |
| $V_2$ | 0.15 (0.25) | 0.13 (0.24) | 0.11 (0.20) | 0.13 (0.23) | 0.13 (0.22) | 0.11 (0.20) | 0.11 (0.17) | 0.11 (0.16) | 0.10 (0.17) |
| | 0.16 (0.28) | 0.13 (0.23) | 0.12 (0.19) | 0.13 (0.23) | 0.13 (0.23) | 0.11 (0.19) | 0.11 (0.17) | 0.10 (0.18) | 0.10 (0.17) |

* Bold values indicate that the 95% confidence intervals of the percentile estimates for boys and girls do not overlap.

**Table 3.4.** Q-wave amplitudes (mV) for boys (upper row) and girls (lower row): median (98th percentile)

| Lead | 0-1 m | 1-3 m | 3-6 m | 6-12 m | 1-3 y | 3-5 y | 5-8 y | 8-12 y | 12-16 y |
|---|---|---|---|---|---|---|---|---|---|
| II | 0.14 (0.23) | 0.18 (0.32) | 0.14 (0.34) | 0.18 (0.48) | 0.15 (0.44) | 0.11 (0.26) | 0.10 (0.28) | 0.09 (0.24) | 0.08 (0.21) |
| | 0.09 (0.26) | 0.14 (0.32) | 0.15 (0.43) | 0.16 (0.44) | 0.16 (0.48) | 0.13 (0.27) | 0.08 (0.26) | 0.08 (0.21) | 0.09 (0.20) |
| III | 0.15 (0.26) | 0.29 (0.50) | 0.31 (0.71) | 0.35 (0.79) | 0.30 (0.74) | 0.19 (0.46) | 0.15 (0.36) | 0.10 (0.28) | 0.10 (0.29) |
| | 0.18 (0.35) | 0.24 (0.50) | 0.28 (0.65) | 0.34 (0.79) | 0.31 (0.73) | 0.18 (0.40) | 0.16 (0.38) | 0.10 (0.27) | 0.10 (0.21) |
| aVF | 0.13 (0.23) | 0.20 (0.35) | 0.20 (0.40) | 0.22 (0.58) | 0.20 (0.54) | 0.14 (0.34) | 0.12 (0.25) | 0.09 (0.25) | 0.08 (0.23) |
| | 0.10 (0.27) | 0.17 (0.35) | 0.20 (0.44) | 0.23 (0.52) | 0.20 (0.54) | 0.12 (0.31) | 0.11 (0.31) | 0.08 (0.21) | 0.09 (0.18) |
| $V_6$ | 0.11 (0.22) | 0.16 (0.31) | 0.17 (0.35) | 0.20 (0.60) | 0.20 (0.56) | 0.15 (0.42) | 0.12 (0.39) | 0.12 (0.43) | 0.11 (**0.43**)* |
| | 0.09 (0.17) | 0.15 (0.37) | 0.15 (0.40) | 0.18 (0.39) | 0.17 (0.49) | 0.15 (0.42) | 0.10 (0.41) | 0.11 (0.34) | 0.09 (**0.23**) |
| $V_7$ | 0.08 (0.13) | 0.13 (0.28) | 0.14 (0.32) | 0.17 (0.52) | 0.19 (0.46) | 0.13 (0.36) | 0.11 (0.30) | 0.11 (0.29) | 0.11 (**0.32**) |
| | 0.08 (0.15) | 0.13 (0.28) | 0.13 (0.36) | 0.16 (0.34) | 0.17 (0.43) | 0.15 (0.33) | 0.09 (0.36) | 0.09 (0.26) | 0.09 (**0.24**) |

* Bold values indicate that the 95% confidence intervals of the percentile estimates for boys and girls do not overlap.

**Table 3.5.** R-wave amplitudes (mV) for boys (upper row) and girls (lower row): median (98th percentile)

| Lead | 0-1 m | 1-3 m | 3-6 m | 6-12 m | 1-3 y | 3-5 y | 5-8 y | 8-12 y | 12-16 y |
|---|---|---|---|---|---|---|---|---|---|
| I | 0.25 (**0.45**)* | 0.56 (1.12) | 0.80 (**1.52**) | 0.82 (1.52) | 0.77 (1.37) | 0.63 (1.09) | **0.62** (**1.16**) | 0.59 (1.04) | **0.58** (1.09) |
| | 0.31 (**0.62**) | 0.55 (1.09) | 0.74 (**1.26**) | 0.75 (1.38) | 0.68 (1.52) | 0.65 (1.20) | **0.49** (**1.00**) | 0.54 (1.21) | **0.48** (1.02) |
| II | 0.64 (1.28) | 1.08 (1.76) | 1.27 (**1.97**) | 1.27 (2.09) | 1.27 (2.47) | 1.36 (2.20) | 1.24 (2.42) | 1.39 (2.23) | 1.31 (2.08) |
| | 0.70 (1.21) | 1.15 (2.04) | 1.33 (**2.24**) | 1.35 (2.21) | 1.27 (2.34) | 1.38 (2.24) | 1.33 (2.27) | 1.32 (2.29) | 1.32 (2.03) |
| III | 0.79 (1.44) | 0.76 (1.60) | 0.72 (**1.50**) | 0.82 (**1.65**) | 0.80 (1.96) | 0.94 (1.82) | 0.80 (1.92) | 0.89 (1.86) | 0.85 (1.74) |
| | 0.85 (1.50) | 0.91 (1.82) | 0.95 (**1.85**) | 0.90 (**1.95**) | 0.96 (2.00) | 0.94 (1.96) | 1.03 (2.09) | 0.92 (1.88) | 0.88 (1.66) |
| aVR | 0.32 (0.52) | **0.36** (**0.63**) | **0.32** (0.58) | **0.30** (**0.62**) | 0.21 (0.53) | 0.21 (**0.48**) | **0.23** (**0.51**) | **0.24** (**0.49**) | **0.23** (**0.46**) |
| | 0.30 (0.61) | **0.27** (**0.49**) | **0.23** (0.51) | **0.21** (**0.48**) | 0.25 (0.48) | 0.17 (**0.39**) | **0.18** (**0.40**) | **0.18** (**0.41**) | **0.18** (**0.37**) |
| aVL | 0.16 (0.32) | **0.35** (0.66) | 0.40 (1.09) | 0.44 (1.04) | 0.38 (0.86) | 0.26 (0.58) | 0.22 (0.70) | 0.17 (0.52) | 0.19 (0.69) |
| | 0.18 (0.45) | **0.25** (0.69) | 0.37 (0.78) | 0.40 (0.92) | 0.38 (1.02) | 0.24 (0.70) | 0.18 (0.55) | 0.17 (0.69) | 0.16 (0.53) |
| aVF | 0.59 (1.36) | 0.88 (1.58) | 0.93 (1.70) | 0.96 (1.81) | 1.00 (2.20) | 1.13 (1.97) | **1.00** (2.19) | 1.16 (2.00) | 1.06 (1.88) |
| | 0.72 (1.26) | 0.98 (1.91) | 1.07 (1.82) | 1.11 (2.04) | 1.10 (2.08) | 1.14 (2.06) | **1.20** (2.17) | 1.09 (2.06) | 1.10 (1.84) |
| $V_{3R}$ | 0.62 (1.04) | 0.58 (**1.24**) | 0.57 (1.20) | 0.48 (1.24) | 0.49 (1.06) | 0.41 (**0.81**) | 0.23 (0.63) | 0.22 (0.51) | 0.19 (0.54) |
| | 0.68 (1.26) | 0.55 (**0.93**) | 0.49 (1.11) | 0.42 (0.98) | 0.43 (0.92) | 0.34 (**0.64**) | 0.21 (0.57) | 0.19 (0.47) | 0.17 (0.49) |
| $V_1$ | 1.10 (2.05) | 1.23 (2.07) | 1.32 (2.20) | 1.12 (2.14) | 1.08 (2.11) | 0.95 (**1.78**) | 0.63 (1.48) | 0.54 (1.14) | **0.48** (1.18) |
| | 1.35 (2.22) | 1.17 (1.99) | 1.14 (2.04) | 1.01 (1.92) | 1.01 (1.91) | 0.77 (**1.38**) | 0.55 (1.24) | 0.49 (1.14) | **0.35** (1.10) |
| $V_2$ | 1.83 (**2.67**) | 1.82 (2.63) | **2.08** (2.54) | 1.94 (**2.51**) | 1.82 (2.41) | 1.58 (2.26) | 1.21 (**2.22**) | **1.02** (1.90) | **0.94** (**1.87**) |
| | 1.83 (**2.17**) | 1.81 (2.45) | **1.88** (2.60) | 1.82 (**2.36**) | 1.75 (2.38) | 1.41 (2.25) | 1.06 (**1.91**) | **0.90** (1.86) | **0.69** (**1.57**) |
| $V_4$ | 1.80 (**2.62**) | 2.30 (3.05) | 2.32 (3.23) | 2.27 (3.32) | 2.37 (3.38) | 2.42 (3.30) | 2.11 (3.11) | 1.86 (3.16) | **1.87** (**3.06**) |
| | 1.68 (**2.21**) | 2.26 (3.26) | 2.26 (3.31) | 2.23 (3.09) | 2.21 (3.54) | 2.24 (3.38) | 1.84 (3.04) | 1.72 (3.23) | **1.24** (**2.55**) |
| $V_6$ | 1.00 (1.78) | 1.55 (**2.23**) | 1.65 (2.73) | 1.70 (2.79) | 1.79 (2.96) | 1.94 (3.14) | 1.97 (2.98) | **2.18** (3.24) | **2.02** (**3.05**) |
| | 0.93 (1.64) | 1.51 (**2.67**) | 1.60 (2.80) | 1.68 (2.74) | 1.68 (2.67) | 1.89 (2.91) | 2.05 (3.25) | **2.00** (3.04) | **1.65** (**2.52**) |
| $V_7$ | 0.45 (0.93) | 0.90 (**1.41**) | 1.01 (1.76) | 1.04 (1.84) | 1.14 (1.99) | 1.34 (2.12) | 1.26 (2.01) | 1.38 (2.24) | 1.41 (**2.31**) |
| | 0.52 (0.96) | 0.95 (**1.68**) | 0.96 (1.80) | 1.13 (1.85) | 1.15 (1.86) | 1.35 (2.12) | 1.36 (2.31) | 1.35 (2.10) | 1.34 (**1.98**) |

* Bold values indicate that the 95% confidence intervals of the percentile estimates for boys and girls do not overlap.

**Table 3.6.** S-wave amplitudes (mV) for boys (upper row) and girls (lower row): median (98th percentile)

| Lead | 0-1 m | 1-3 m | 3-6 m | 6-12 m | 1-3 y | 3-5 y | 5-8 y | 8-12 y | 12-16 y |
|---|---|---|---|---|---|---|---|---|---|
| I | 0.42 (0.71) | **0.46** (**0.94**) | **0.41** (0.77) | **0.40** (**0.81**) | 0.27 (0.82) | 0.21 (0.69) | 0.22 (0.56) | **0.22** (0.50) | **0.19** (0.48) |
| | 0.51 (1.01) | **0.35** (**0.71**) | **0.32** (0.73) | **0.33** (**0.73**) | 0.35 (0.70) | 0.20 (0.52) | 0.22 (0.54) | **0.16** (0.47) | **0.13** (0.40) |
| II | 0.24 (0.46) | 0.29 (0.55) | 0.29 (**0.61**) | **0.30** (0.62) | 0.25 (0.55) | **0.28** (**0.58**) | **0.27** (**0.64**) | **0.30** (**0.63**) | 0.27 (**0.63**) |
| | 0.26 (0.53) | 0.22 (0.53) | 0.24 (**0.46**) | **0.23** (0.54) | 0.26 (0.56) | **0.20** (**0.46**) | **0.19** (**0.46**) | **0.20** (**0.52**) | 0.22 (**0.54**) |
| III | 0.16 (**0.28**) | 0.27 (0.54) | 0.30 (**0.87**) | 0.34 (0.86) | 0.30 (0.72) | 0.22 (0.51) | 0.21 (**0.65**) | 0.19 (0.56) | 0.20 (0.57) |
| | 0.19 (**0.34**) | 0.24 (0.50) | 0.28 (**0.63**) | 0.33 (0.77) | 0.32 (0.86) | 0.19 (0.54) | 0.18 (**0.41**) | 0.16 (0.48) | 0.17 (0.61) |
| aVR | 0.41 (0.68) | 0.76 (1.30) | 0.98 (1.47) | 0.98 (1.47) | 0.95 (1.63) | 0.93 (1.40) | 0.90 (1.51) | 0.96 (1.45) | 0.91 (1.39) |
| | 0.44 (0.64) | 0.81 (1.31) | 0.96 (1.49) | 0.97 (1.48) | 0.92 (1.61) | 0.95 (1.49) | 0.90 (1.40) | 0.91 (1.51) | 0.89 (1.35) |
| aVL | 0.47 (**0.77**) | 0.51 (1.02) | 0.44 (**0.83**) | 0.47 (0.98) | 0.40 (1.00) | 0.34 (0.87) | 0.33 (**0.84**) | 0.28 (0.88) | 0.28 (0.94) |
| | 0.63 (**1.17**) | 0.53 (1.04) | 0.46 (**0.98**) | 0.52 (1.03) | 0.44 (1.06) | 0.33 (1.12) | 0.43 (**1.02**) | 0.30 (0.88) | 0.28 (0.84) |
| aVF | 0.18 (**0.27**) | 0.22 (0.39) | 0.23 (0.57) | 0.23 (0.59) | 0.23 (0.53) | **0.22** (0.52) | 0.21 (**0.57**) | 0.21 (**0.56**) | 0.22 (0.54) |
| | 0.18 (**0.38**) | 0.20 (0.35) | 0.20 (0.44) | 0.24 (0.51) | 0.24 (0.60) | **0.16** (0.40) | 0.16 (**0.37**) | 0.17 (**0.45**) | 0.18 (0.55) |
| $V_{3R}$ | 0.12 (0.22) | **0.24** (0.86) | 0.31 (0.90) | 0.34 (1.04) | 0.45 (1.21) | 0.53 (0.99) | 0.53 (1.06) | 0.60 (1.17) | 0.57 (1.14) |
| | 0.25 (0.62) | **0.35** (0.76) | 0.31 (0.98) | 0.34 (0.95) | 0.42 (1.08) | 0.50 (1.16) | 0.52 (1.07) | 0.55 (1.20) | 0.50 (1.04) |
| $V_1$ | 0.74 (1.41) | 0.63 (1.57) | 0.69 (2.02) | 0.69 (1.88) | 0.95 (2.27) | 1.09 (2.11) | 1.15 (2.29) | 1.30 (2.46) | 1.30 (**2.44**) |
| | 0.72 (1.48) | 0.82 (1.59) | 0.74 (1.64) | 0.76 (1.86) | 0.86 (2.13) | 1.03 (2.11) | 1.23 (2.49) | 1.32 (2.58) | 1.15 (**2.05**) |
| $V_2$ | 1.53 (2.40) | 1.26 (2.54) | 1.49 (2.48) | 1.50 (2.78) | 1.77 (2.95) | 2.01 (3.08) | 2.17 (3.25) | 2.28 (3.44) | **2.39** (**3.58**) |
| | 1.47 (2.47) | 1.55 (2.61) | 1.47 (2.48) | 1.56 (2.52) | 1.70 (2.91) | 1.96 (2.93) | 2.17 (3.49) | 2.29 (3.46) | **1.87** (**3.14**) |
| $V_4$ | 1.17 (1.71) | 1.11 (2.25) | 1.22 (2.42) | **1.25** (2.35) | **1.16** (2.16) | **1.25** (**2.51**) | **1.28** (**2.68**) | **1.31** (2.44) | **1.16** (**2.23**) |
| | 1.04 (1.87) | 1.18 (1.87) | 1.19 (2.18) | **0.98** (2.04) | **0.91** (2.00) | **0.97** (**1.75**) | **1.05** (**2.33**) | **1.00** (2.28) | **0.73** (**1.60**) |
| $V_6$ | 0.49 (0.77) | 0.51 (**1.12**) | 0.46 (1.25) | **0.46** (**1.21**) | 0.37 (0.91) | 0.34 (**0.86**) | **0.34** (0.89) | **0.34** (0.79) | **0.37** (**0.85**) |
| | 0.44 (1.07) | 0.39 (**0.77**) | 0.41 (0.97) | **0.31** (**0.70**) | 0.33 (0.88) | 0.30 (**0.61**) | **0.29** (0.77) | **0.27** (0.75) | **0.30** (**0.67**) |
| $V_7$ | 0.18 (0.31) | **0.24** (0.46) | 0.22 (0.50) | **0.26** (**0.58**) | 0.22 (0.53) | 0.21 (0.41) | 0.17 (0.39) | **0.16** (0.39) | **0.20** (**0.38**) |
| | 0.16 (0.37) | **0.18** (0.39) | 0.19 (0.43) | **0.20** (**0.37**) | 0.21 (0.48) | 0.17 (0.36) | 0.13 (0.40) | **0.12** (0.33) | **0.16** (**0.34**) |

* Bold values indicate that the 95% confidence intervals of the percentile estimates for boys and girls do not overlap.

**Table 3.7.** R/S ratio in precordial leads for boys (upper row) and girls (lower row): median (2nd percentile, 98th percentile)

| Lead | 0-1 m | 1-3 m | 3-6 m | 6-12 m | 1-3 y | 3-5 y | 5-8 y | 8-12 y | 12-16 y |
|---|---|---|---|---|---|---|---|---|---|
| $V_{3R}$ | 2.4 (1.2,S=0)* | 2.4 (0.5,S=0) | 1.8 (0.4,S=0) | 1.5 (0.4,S=0) | 1.2 (0.4,S=0) | 0.7 (0.1,S=0) | 0.4 (0.0,1.6) | 0.4 (0.1,1.2) | 0.3 (0.1,1.1) |
| | 3.5 (0.0,S=0) | 1.7 (0.5,S=0) | 1.7 (0.2,S=0) | 1.2 (0.2,S=0) | 1.2 (0.2,S=0) | 0.6 (0.1,S=0) | 0.4 (0.1,1.7) | 0.3 (0.0,1.2) | 0.3 (0.0,1.5) |
| $V_1$ | 1.6 (0.8,3.7) | 1.9 (0.5,5.0) | 1.8 (0.4,4.9) | 1.6 (0.7,4.2) | 1.2 (0.5,2.9) | 0.8 (0.3,1.9) | 0.6 (0.1,1.7) | 0.4 (0.1,1.2) | 0.4 (0.1,1.1) |
| | 1.8 (1.0,4.9) | 1.5 (0.6,4.4) | 1.6 (0.4,4.1) | 1.4 (0.4,3.4) | 1.2 (0.5,2.8) | 0.7 (0.2,1.8) | 0.5 (0.1,1.4) | 0.4 (0.1,1.1) | 0.3 (0.1,1.0) |
| $V_2$ | 1.1 (0.7,2.3) | 1.4 (0.6,2.8) | 1.3 (0.8,2.2) | 1.3 (0.7,2.5) | 1.0 (0.5,2.4) | 0.8 (0.3,1.5) | 0.5 (0.1,1.3) | 0.5 (0.1,1.1) | 0.4 (0.1,1.1) |
| | 1.3 (0.7,2.5) | 1.1 (0.7,2.8) | 1.3 (0.6,2.9) | 1.2 (0.5,2.2) | 1.1 (0.4,1.7) | 0.7 (0.3,1.5) | 0.5 (0.1,1.2) | 0.4 (0.1,1.2) | 0.4 (0.1,1.0) |
| $V_6$ | 1.9 (1.0,3.7) | 3.0 (0.8,8.3) | 3.6 (0.4,S=0) | 3.7 (1.1,S=0) | 5.0 (0.8,S=0) | 6.1 (1.9,S=0) | 5.9 (1.8,S=0) | 6.2 (1.7,S=0) | 5.5 (2.0,S=0) |
| | 2.0 (1.0,3.7) | 3.6 (1.7,8.7) | 4.0 (1.1,S=0) | 4.9 (1.8,S=0) | 5.6 (0.5,S=0) | 7.2 (2.7,S=0) | 6.8 (1.7,S=0) | 7.2 (2.0,S=0) | 5.4 (1.3,S=0) |
| $V_7$ | 2.9 (1.2,S=0) | 3.6 (1.4,S=0) | 4.4 (1.1,S=0) | 4.8 (1.4,S=0) | 6.8 (2.0,S=0) | 8.1 (2.0,S=0) | 8.2 (2.2,S=0) | 8.4 (1.6,S=0) | 7.4 (3.1,S=0) |
| | 2.8 (0.7,S=0) | 4.9 (1.6,S=0) | 5.3 (2.2,S=0) | 6.3 (2.1,S=0) | 6.3 (0.6,S=0) | 8.1 (1.8,S=0) | 8.3 (2.3,S=0) | 10.0 (3.0,S=0) | 7.6 (2.7,S=0) |

* S waves were absent in more than two percent of the ECGs (S=0).

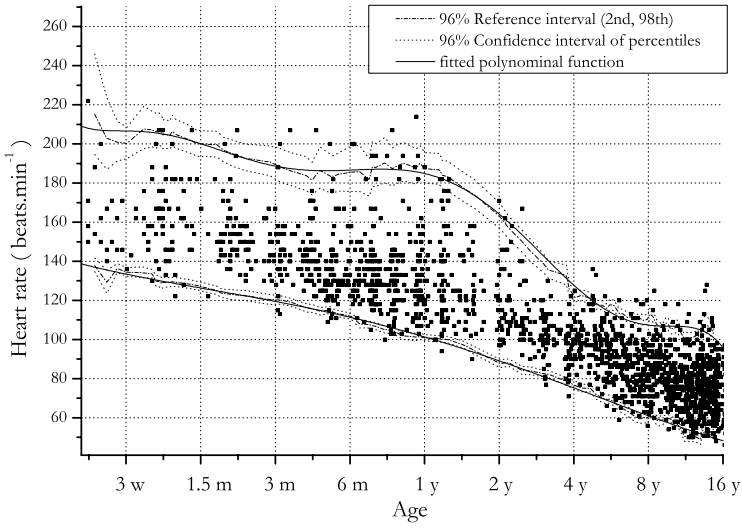


**Figure 3.1.** Continuous age-dependent percentile curves of the heart rate for the total population.

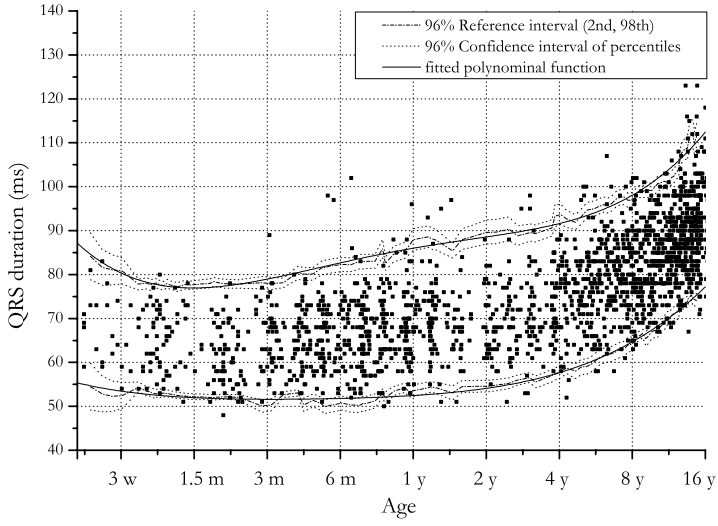


**Figure 3.2.** Continuous age-dependent percentile curves of the QRS duration for the total population.

In Table 3.7, the R/S ratio is presented for the precordial leads. Although a steady decrease is observed, the median R/S ratio in $V_1$ remains greater than one up to three years of age. In some age groups the ULN could not be calculated because S waves were absent in more than two percent of the ECGs.

## 3.4 Discussion

ECGs of healthy children change markedly from birth to young adulthood. Knowledge of the normal variation of ECG measurements with age is essential for proper interpretation of the pediatric ECG. Previous studies that determined normal limits for the pediatric ECG had their imperfections: ECGs were recorded with relatively low sampling rate, ECG measurements were done manually, or normal limits were presented for only a limited set of parameters. In this study, normal limits of ECG parameters were based on computerized analysis of a large set of ECGs recorded at a high sampling rate, thus obviating some of the limitations of previous studies. These new normal limits differ substantially from the limits presented by Davignon et al. [4], which are commonly used in pediatric electrocardiography, and may call for changes in diagnostic ECG criteria. We will discuss some implications for the assessment of prolonged QRS and QTc-interval duration, and for the diagnosis of right atrial hypertrophy and ventricular hypertrophy.

Normal limits for the QRS duration are substantially higher than those reported by Davignon et al. [4]. For instance, children aged 12-16 years had a median QRS duration of 90 ms compared to 65 ms in the Davignon study. However, Davignon only calculated the QRS duration in $V_5$, whereas MEANS determines the QRS duration over all leads, which yields longer QRS durations. Our findings corroborate with the study of Macfarlane et al. [6], who reported a mean QRS duration of 86 ms for children aged 13-14 year.

QTc-interval prolongation is a valuable tool for detecting and quantifying the risk of arrhythmia due to drugs [21,22]. Moreover, QTc interval prolongation has been associated with sudden infant death syndrome or apparently life threatening incidents in infants [23]. Valid normal values are a prerequisite for proper interpretation in these studies. We found an ULN for the QTc interval of approximately 450 ms, which is higher than the commonly used criterion of 440 ms [24].

For the diagnosis of right atrial hypertrophy (RAH), the P-wave amplitude should be greater than 0.25 mV [24] or 0.30 mV [2,5] in any lead. This criterion is based on the upper limit of the normal P-wave amplitude. In our study the ULN of the P-wave

amplitude is 0.25 mV in lead II, while in $V_1$ and $V_2$ substantially lower ULNs are found (Table 3.3). These results suggest that the amplitude criterion in the diagnosis of RAH should be made lead dependent.

In diagnosing ventricular hypertrophy, amplitude criteria for different ECG parameters are employed. Deep Q waves in $V_6$ are suggestive of left ventricular hypertrophy (LVH) [5]. The ULN of the Q-wave amplitude in our study is substantially higher than presented by Davignon et al. [4]. For example, for children aged 3-5 years we found an ULN of the Q-wave amplitude of 0.54 mV against 0.30 mV in the Davignon study. Macfarlane et al. [6] obtained similar results for Q-wave amplitudes in neonates. Considering that narrow deep Q waves contain relatively high frequencies, our findings may demonstrate the effect of using a higher sampling rate. Another reason that may partly explain the differences is that we only included non-zero values in computing the percentiles. It is not clear whether this was also done in the Davignon study. When we recomputed the ULN of the Q-wave amplitude with zero values included, the ULN decreased to 0.47 mV. However, because a Q wave is defined as a negative deflection, we believe the exclusion of zero values is the preferred approach.

R-wave and S-wave amplitudes in the precordial leads are important parameters in the diagnosis of both right and left ventricular hypertrophy. We found considerable differences in R- and S-wave amplitudes compared to Davignon et al. [4], especially in $V_6$. For example, in our study the median of the R-wave amplitude in $V_6$ for children aged 8-12 years is 2.09 mV as compared to 1.68 mV in the study of Davignon. Higher R-wave amplitudes in $V_6$ were also presented by Macfarlane et al. [6], who found a mean R-wave amplitude in $V_6$ of 1.9 mV for children aged 5-12 years. For all age groups, the ULNs of the R-wave amplitude in $V_6$ are substantially higher in our study, e.g., 3.14 mV for children aged 5-8 years compared to 2.65 mV in the study of Davignon. Notably, the ULN of the R-wave amplitude in $V_{3R}$, $V_2$ and especially $V_4$ is lower in almost all age groups. For instance, Davignon reports an ULN of 4.5 mV in $V_4$ for children aged 3-5 year, compared to 3.27 mV in our study. R-wave amplitudes in $V_4$ larger than 3.5 mV are exceptional in our material. S-wave amplitudes are considerably larger than reported by Davignon in $V_6$ for all age groups, and in $V_4$ after 3 years of age. In the other precordial leads the S-wave amplitude is comparable in most age groups. These findings suggest that the amplitude criteria for ventricular hypertrophy should be adjusted.

Influence of sex differences on the pediatric ECG has been reported in a number of studies [3,7,8,25–27]. However, to our knowledge this is the first major study that examined sex differences in amplitude measurements for children in all age groups. In our study, amplitudes of the Q, R and S waves are higher for boys than for girls during

adolescence in most precordial leads. For example, the ULN of the R wave in $V_6$ is 3.05 mV for boys and 2.55 mV for girls in the age group of 12 to 16 years. Little change in voltages is seen in boys during adolescence, while in girls a progressive decline is observed. In a study of 114 adolescents, Strong et al. [3] stated that the sex differences were primarily a reflection of the boys being greater than girls of reproductive age. Another reason for the amplitude differences during adolescence could be the development of breast tissue [28]. Moreover, we found clinically significant differences at younger ages, especially in the S waves in the left precordial leads. At this point we do not have any further elucidation for the reason of these sex differences. The amplitude differences are substantial and indicate that sex-dependent criteria could improve the sensitivity and specificity for left ventricular hypertrophy in children. For adolescents, this has already been noted in the early seventies by Walker et al. [25], but is in our experience not used in daily practice. Furthermore, effects of sex on ECG interval measurements were seen for QRS duration, which is consistently longer for boys in all age groups. This was also previously shown by Macfarlane et al. [7]. No substantial sex differences for the QTc interval could be demonstrated. However, in the group of 12-16 years the confidence intervals of the ULN of the QTc interval only marginally overlapped, indicating possibly longer QTc intervals for girls. In a recent study, Eberle et al. [27] also suggested that there is an influence of gender on the QT interval in the group of 13 to 16 years. Pearl et al. [8] demonstrated significantly longer QTc intervals for girls from the age of 14 years. The difference appears to be due to QT shortening in boys rather than a QT prolongation in girls [26].

A minimal sampling rate of 500 Hz has been recommended for the adult ECG [29], but for pediatric ECGs higher sampling rates have been suggested [11,12]. We used a sampling rate of 1200 Hz, which was deemed sufficiently high to accurately record pediatric ECGs. When we downsampled the signals to 500 Hz and repeated our analyses, normal limits remained essentially the same. However, when we downsampled to 333 Hz as used by Davignon [4], lower amplitudes were found, e.g., R-wave amplitudes in $V_6$ decreased up to 0.15 mV. Still, we consider it unlikely that amplitude differences between our study and Davignon's can solely be attributed to differences in sampling rate, since we found lower QRS amplitudes in some leads. Other factors may also play a role, such as population differences and physiological changes in children, e.g., length, in the twenty years that passed between both studies.

We chose to present most of our results in tables rather than in plots because of space limitations. However, one should be aware that the tabulated normal values are estimates for the median age in the age groups and that an age-effect within age groups may still be present. For children with ages close to the boundary of an age group, it is prudent to interpolate normal values between adjacent age groups. This is well illus-

trated by the continuous age-dependent percentile curves of the heart rate in Fig. 3.1, which shows a strong age-dependency within the age group of one to three years. Moreover, continuous age-dependent curves are preferred for computerized interpretation of pediatric ECGs, since they help to avoid abrupt changes in diagnosis with small differences in age.

## 3.5 Study Limitations

The normal limits of the group of 0-1 month should be used with caution, because the sample size of this group is relatively small and our database does not contain ECGs recorded during the first ten days after birth. The collection of the huge number of ECGs, necessary to obtain reliable estimates of normal limits for the youngest ages, would require a further study.

## 3.6 Conclusions

Normal limits have been estimated for pediatric ECGs recorded at a high sampling rate of 1200 Hz and analysed with the use of a computer program, thus obviating some of the limitations of previous studies. Normal limits of many ECG measurements were shown to differ from those reported earlier. Significant sex differences could be demonstrated for amplitude measurements and QRS duration. These findings are clinically significant and suggest that diagnostic criteria for the pediatric ECG should be adjusted.

## 3.7 Acknowledgments

The authors wish to thank Joke van Woerkom for recording all ECGs, and the children for their participation in our study. This study was supported by the Netherlands Ministry of Economic Affairs (Senter grant ITU94035).

## 3.8 References

[1] Maroney M, Rantz LA. Electrocardiograms in 679 healthy infants and children. Pediatrics 1950;5:396.

[2] Ziegler RF. Electrocardiographic studies in normal infants and children. Springfield: Charles C. Thomas, 1951.

[3] Strong WB, Downs TD, Liebman J, Liebowitz R. The normal adolescent electrocardiogram. Am Heart J 1972;83:115–28.

[4] Davignon A, Rautaharju P, Boisselle E, Soumis F, Megelas M, Choguette A. Normal ECG standards for infants and children. Pediatr Cardiol 1979/80;1:123–31.

[5] Liebman J, Plonsey R, Gillette PC. Pediatric cardiology. Baltimore: Williams & Wilkins, 1982.

[6] Macfarlane PW, Coleman EN, Pomphrey EO, McLaughlin S, Houston A, Aitchison T. Normal limits of the high-fidelity pediatric ECG. Preliminary observations. J Electrocardiol 1989;22:162–8.

[7] Macfarlane PW, McLaughlin SC, Devine B, Yang TF. Effects of age, sex, and race on ECG interval measurements. J Electrocardiol 1994;27:14–9.

[8] Pearl W. Effects of gender, age, and heart rate on QT intervals in children. Pediatr Cardiol 1996;17:135–6.

[9] Tutar HE, Ocal B, Imamoglu A, Atalay S. Dispersion of QT and QTc interval in healthy children, and effects of sinus arrhythmia on QT dispersion. Heart 1998;80:77–9.

[10] Golden DP, Wolthuis RA, Hoffler GW. A spectral analysis of the normal resting electrocardiogram. IEEE Trans Biomed Eng 1973;20:366–72.

[11] Barr RC, Spach MS. Sampling rates required for digital recording of intracellular and extracellular cardiac potentials. Circulation 1977;55:40–48.

[12] Garson A. J. Clinically significant differences between the "old" analog and the "new" digital electrocardiograms. Am Heart J 1987;114:194–7.

[13] Van Wieringen JC, Roede MJ, Wit JM. Groeidiagrammen voor patientenzorg [Growth diagrams for patient care]. Tijdschrift Kindergeneeskunde 1985;53:147–52.

[14] Van Bemmel JH, Kors JA, Van Herpen G. Methodology of the modular ECG analysis system MEANS. Methods Inf Med 1990;29:346–53.

[15] Willems JL, Arnaud P, Van Bemmel JH, Bourdillon PJ, Degani R, Denis B, et al. A reference data base for multilead electrocardiographic computer measurement programs. J Am Coll Cardiol 1987;10:1313–21.

[16] Willems JL, Zywietz C, Arnaud P, van Bemmel JH, Degani R, Macfarlane PW. Influence of noise on wave boundary recognition by ECG measurement programs. Recommendations for preprocessing. Comput Biomed Res 1987;20:543–62.

[17] Solberg HE. Approved recommendation (1986) on the theory of reference values. Part 1. The concept of reference values. J Clin Chem Clin Biochem 1987;25:337–342.

[18] Manly BFJ. Exponential data transformations. The statistician 1976;25:37–42.

[19] John JA, Draper NR. An alternative family of transformations. Applied Statistics 1980; 29:190–7.

[20] Bazett H. An analysis of the time-relations of electrocardiograms. Heart 1918;7:353–70.

[21] Moss AJ. Measurement of the QT interval and the risk associated with QTc interval prolongation: a review. Am J Cardiol 1993;72:23B–25B.

[22] Hill SL, Evangelista JK. Proarrhythmia associated with casapride in children. Pediatrics 1998;101:1053–6.

[23] Schwartz PJ, Stramba-Badiale M. Prolongation of the QT interval and the sudden infant death syndrome. N Engl J Med 1998;338:1709–14.

[24] Park MK, Guntheroth G. How to read pediatric ECGs. St Louis: Mosby-Year Book Inc., 1992.

[25] Walker CHM, Rose RL. Importance of age, sex and body habitus in the diagnosis of left ventricular hypertrophy from the precordial electrocardiogram in childhood and adolescence. Pediatrics 1961;28:705.

[26] Rautaharju PM, Zhou SH, Wong S, Calhoun HP, Berenson GS, Prineas R, et al. Sex differences in the evolution of the electrocardiographic QT interval with age. Can J Cardiol 1992;8:690–5.

[27] Eberle T, Hessling G, Ulmer HE, Brockmeier K. Prediction of normal QT intervals in children. J Electrocardiol 1998;31 (Suppl):121–5.

[28] LaMonte CS, Freiman AH. The electrocardiogram after mastectomy. Circulation 1965; 32:746–54.

[29] Pipberger HV, Arzbaecher RC, Berson AS, Brody DA, Flowers C, Geselowitz DB, et al. Recommendations for standardization of leads and specifications for instruments in electrocardiography and vectorcardiography. Report of the Committee on Electrocardiography, American Heart Association. Circulation 1975;52:11–31.

# 4 Exhaustive rule induction under constraints

P.R. Rijnbeek, J.A. Kors

Dept. of Medical Informatics, Erasmus University Medical Center, Rotterdam, The Netherlands

Submitted for publication

## Abstract

Most learning algorithms are not evaluating every description which may result in a suboptimal classifier. Furthermore, they are aimed at the maximization of the accuracy of the classifier. However, in many application areas, for example in medicine, sensitivity and specificity are more pertinent measures to characterize classification performance. In this study, we present a new version of our induction algorithm, called EXPLORE (Exhaustive Procedure for LOgic-Rule Extraction), that exhaustively generates rules that fulfill user-specified performance requirements. Several new techniques are introduced to search the feature space more efficiently than by a straightforward, brute-force exhaustive search. The new version of EXPLORE, incorporating these techniques, is able to perform an exhaustive search considerably faster. On five standard data sets from the medical domain the accuracy of the best rule generated by EXPLORE is comparable to or surpasses the accuracy of the classifiers generated by the greedy algorithms C4.5 and CART.


**Keywords:** learning, induction, exhaustive, algorithm

## 4.1 Introduction

There exist many algorithms by which classifiers, such as decision rules or decision trees are learned from a set of example cases [1–5]. Each case consists of a set of feature values and a label indicating the class to which the case belongs. These induction algorithms search the feature space to generate the best classifier in terms of the proportion of correctly classified cases. To find this best classifier, most induction algorithms use a "greedy" approach to search the feature space, i.e., they do not consider all possible classifiers but build up a classifier iteratively. At each iteration the feature test that performs best according to some criterion is selected, without reconsidering the selections that were made during the previous steps. Greedy algorithms have been shown to perform well on a variety of problems, but they also have several limitations.

In the first place, greedy approaches may miss the best classifier. Only an exhaustive search of the feature space guarantees to find this classifier, so that a greedy approach would only be called for if an exhaustive search is not feasible. Unfortunately this is often the case, even with ever-increasing computing power and exhaustive search is still unwieldy for all but the simplest classification problems. A second limitation of greedy algorithms is that they are aimed at the maximization of the accuracy of the classifier that is being learned. However, in many application areas, for example in medicine, sensitivity and specificity are more pertinent measures to characterize classification performance [6]. A classifier in these areas often needs to be maximal for either one of these performance measures while fulfilling a user-specified minimum performance for the other. Such optimization under constraints is difficult to achieve with greedy algorithms by the nature of their method to select feature tests, but is straightforward when an exhaustive search algorithm is used.

In the past, a few algorithms have been developed that approximate an exhaustive search [7–9]. The Predictive Value Maximization (PVM) algorithm described by Weiss et al. [7] generates decision rules based on a modified beam search strategy. Beam search is a heuristic search algorithm that is an optimization of the greedy approach. Like the greedy approach, it uses a heuristic function to estimate the promise of each node it examines. Beam search, however, unfolds a fixed number of most promising nodes at each depth. The beam search heuristic is a massive search, but it is not exhaustive because it does not generate all possible combinations of feature tests. Nevertheless, Weiss showed that this massive search strategy often outperforms a greedy search. An attractive feature of the PVM algorithm is that it enables the desirable "optimization under constraints". Another massive search algorithm, called BRUTE, was also based on a beam search approach [8]. In BRUTE the search space is reduced

to manageable proportions without loss of performance using an admissable search strategy [9]. BRUTE was shown to perform better than greedy search algorithms on a large number of data sets [8]. A disadvantage of BRUTE is that it can only generate rules consisting of conjunctions of feature tests, whereas rules that allow both conjunctions and disjunctions may be more appropriate in many situations. The algorithm allows the user to specify a performance measure to be optimized, but it is not possible to simultaneously impose constraints on other performance measures.

To our knowledge, two algorithms that perform a truly exhaustive search have been described. The 1R algorithm [10] searches exhaustively for the best classification rule of length one, that is, a rule that contains only one feature test. Such very simple classification rules were shown to perform surprisingly well on many data sets, but obviously, longer rules that could perform much better are missed. Previously, we ourselves developed an induction algorithm, called EXPLORE (Exhaustive Procedure for LOgic-Rule Extraction), which exhaustively generates rules that fulfill user-specified performance requirements [11]. In this earlier version of EXPLORE the rules were generated exhaustively using a time-consuming brute-force approach, which limited its practical applicability.

In this study, we present several new techniques to search the feature space more efficiently than a straightforward, brute-force exhaustive search. We show that a new version of EXPLORE, incorporating these techniques, is able to perform an exhaustive search considerably faster.

## 4.2 Exhaustive rule induction under constraints

The EXPLORE induction algorithm generates the best decision rule for a two-class classification problem. The user defines which performance measure should be optimized and specifies constraints on other performance measures. Table 4.1 shows the allowable performance measures and illustrates how they can be estimated from a $2 \times 2$ classification matrix.

The algorithm systematically generates all possible rules of a certain length, based on the following format:

if <logical expression> then <class 1> else <class 2>

The logical expression consists of an ensemble of feature tests joined by the logical operators $AND$ and $OR$. Each feature test compares a feature value with a threshold

**Table 4.1.** Estimation of performance measures from a $2 \times 2$ classification matrix.

| | Rule positive | Rule negative |
|---|---|---|
| Class 1 | Correct positives $(CP)$ | False negatives $(FN)$ |
| Class 2 | False positives $(FP)$ | Correct negatives $(CN)$ |

Sensitivity = $CP/(CP + FN)$
Specificity = $CN/(FP + CN)$
Positive predictive value = $CP/(CP + FP)$
Negative predictive value = $CN/(FN + CN)$
Accuracy = $(CP + CN)/(CP + FN + FP + CN)$

value through a relational operator ($\leq$ or $>$). A feature test comes out true or false, dependent on the feature value of a particular cases. The then- and the else-clauses of the rule specify the class to which a case is assigned if the logical expression evaluates to true or false, respectively. EXPLORE generates these logical expressions in Disjunctive Normal Form (DNF), defined as disjunctions ($OR$ ) of conjunctions ($AND$) of feature tests. The length of a rule, $n$, is defined as the number of feature tests in the rule. The ensemble of features and operators in a rule we call a "feature-operator set" and the ensemble of threshold values a "threshold set".

**Brute Force approach**

In the earlier version of the EXPLORE algorithm the rules were generated in a brute-force manner. The number of rules that are generated and evaluated when using this approach is very large and depends on the number of features, operators, and threshold values.

Suppose that each case in the data set is composed of $f$ features and that each feature has $t$ potential threshold values. Further let us assume that rules of length $n$ are to be generated. When only one logical operator, either $AND$ or $OR$ , is allowed $f^n$ different expressions can be made. These expressions can then be instantiated in $2^n$ different ways by the relational operators $\leq$ and $>$, and in $t^n$ ways by the thresholds. When both logical operators are used to generate the expressions, the factor $f^n$ must be multiplied by the $n^{th}$ Bell number, $B_n$, which indicates the number of ways to partition a set of $n$ elements into a set of disjunct subsets [12].

The Bell number is recursively defined as:

$$B_{n+1} = \sum_{k=0}^{n} \binom{n}{k} B_k \tag{4.1}$$

with $B_0 = 1$. For $n = 0, 1, 2, 3$, $B_{n+1} = 1, 2, 5, 15$. The number of rules to generate is therefore:

$$f^n \times 2^n \times t^n \times B_n \tag{4.2}$$

However, the number of threshold values per feature is not fixed and will vary from feature to feature. This can be taken into account by combining the terms $f^n$ and $t^n$. Assuming that feature $k$ has $t_k$ threshold values, one feature test, i.e., a rule of length 1, can be instantiated in $\sum_k t_k$ different ways; similarly, a rule of length $n$ in $(\sum_k t_k)^n$ ways. Thus the number of rules to generate is equal to:

$$\left( \sum_{k=1}^{f} t_k \right)^n \times 2^n \times B_n \tag{4.3}$$

We will call this kind of search a naïve exhaustive search, because in this approach many logically identical rules are generated.

**Systematic rule generation**

EXPLORE consists of three nested do-while loops (Alg. 4.1), which systematically generate all possible and relevant decision rules. The outer loop instantiates a rule of given rule length with logical operators (a logical-operator set), the second loop adds features and relational operators (a feature-operator set), and the inner loop adds threshold values (a threshold set). The fully instantiated rule is then evaluated, and if it fulfills the performance constraints and outperforms the current best rule, the latter is replaced by the new rule. In the following we will give a detailed description of each of the three loops, focussing on the routines that perform loop initialization (InitLogicalOperatorSet, InitFeatureOperatorSet, InitThresholdSet) and loop control (NextLogicalOperatorSet, NextFeatureOperatorSet, NextThresholdSet).

**Algorithm 4.1**: EXPLORE

```
1  InitLogicalOperatorSet;
2  do
3      InitFeatureOperatorSet;
4      do
5          InitThresholdSet;
6          do
7              Calculate performance measures of the current rule;
8              if current rule outperforms best rule AND current rule fulfills constraints then
9                  Replace best rule by current rule;
10             end
11         while NextThresholdSet ;
12     while NextFeatureOperatorSet ;
13 while NextLogicalOperatorSet ;
```

### 4.2.1 Generation of logical-operator sets

The generation of logical-operator sets is implemented as a routine that exhaustively generates the next unique combination of disjunctions and conjunctions. As an example, Table 4.2 shows all unique logical-operator sets for rule length $n = 5$.

**Table 4.2.** Enumeration of all logical-operator sets for rule length $n = 5$. Numbers indicate the number of feature tests in a conjunction.

| |
|---|
| 5 |
| 4 *OR* 1 |
| 3 *OR* 2 |
| 3 *OR* 1 *OR* 1 |
| 2 *OR* 2 *OR* 1 |
| 2 *OR* 1 *OR* 1 *OR* 1 |
| 1 *OR* 1 *OR* 1 *OR* 1 *OR* 1 |

To generate this sequence we reformulate the problem in terms of the partitioning of a positive integer $n$ [13]. A partition of a positive integer $n$ is a finite decreasing sequence of positive integers $(\lambda_1, \lambda_2, \ldots, \lambda_m)$ such that $\sum_{i=1}^{m} \lambda_i = n$. The $\lambda_i$ are called the parts of the integer. In our context of rule generation, a rule of length $n$ is denoted as $\lambda_1$ *OR* $\lambda_2$ *OR* ... *OR* $\lambda_m$, in which $\lambda_i$ indicates the size of conjunction $i$, i.e., the number of feature tests in the conjunction; $\lambda_i = 1$ represents a single feature test. For example, 4 *OR* 1 represents a disjunction of a conjunction of size 4 and a single feature test.

### Initialization of logical-operator set

The initialization of the logical-operator set is extremely simple: we start with a single conjunction with a size that equals the rule length (Alg. 4.2).

**Algorithm 4.2**: InitLogicalOperatorSet

```
1 m = 1;                                  // number of conjunctions is one
2 λ_1 = n;                // the size of the conjunction equals the rule length
```

### Generation of next logical-operator set

To generate the next integer partition, or in our context the next logical-operator set, several algorithms have been developed [14]. They are all based on the general idea of subtracting one from the smallest part $\lambda_i > 1$ and collecting $\lambda_i = 1$ parts to form a next partition. We define $h$ as the number of parts greater than 1, i.e. $\lambda_i > 1$ for $1 \leq i \leq h$, and $\lambda_i = 1$ for $h < i \leq m$. The last integer partitioning consists of $n$ parts of size $\lambda_i = 1$ (Alg. 4.3).

**Algorithm 4.3**: NextLogicalOperatorSet

```
   // Is there a next logical-operator set?
1  if λ_1 > 1 then
2      if λ_m = 1 then
3          Replace λ_h, λ_{h+1} = 1, ..., λ_m = 1 by c − 1 integers equal to (λ_h − 1), and a single integer d,
           such that 0 < d ≤ λ_h − 1 and (λ_h − 1)(c − 1) + d = λ_h + m − h
4      else
5          Replace λ_1, ..., λ_m by λ_1, ..., λ_m − 1, 1
6      end
7      Return TRUE;
8  else
9      Return FALSE
10 end
```

For example, given the logical-operator set 3 $OR$ 1 $OR$ 1 $(m = 3,\ h = 1,\ \lambda_h = 3)$, the next logical-operator set is 2 $OR$ 2 $OR$ 1 $(c = 3,\ d = 1)$.

**Number of logical-operator sets**

Let $p(n,k)$ be the number of partitions with all parts $\lambda_i \geq k$. The number of logical-operator sets, $p(n,1)$, can then be calculated by the following recursive formula [13]:

$$\begin{aligned} p(n,k) &= p(n,k+1) + p(n-k,k) & \\ p(n,k) &= 1 & \text{if } (k=n) \\ p(n,k) &= 0 & \text{if } (k>n) \end{aligned} \tag{4.4}$$

For $n = 1, 2, 3, 4, 5$,p(n,1) = 1, 2, 3, 5, 7. For example, a rule of length $n = 5$ will have 7 different logical-operator set (Table 4.2).

### 4.2.2 Generation of feature-operator sets

After a new logical-operator set has been generated, the rule will be instantiated with features $(A, B, C, \ldots)$ and operators $(\leq, >)$. Several decision rule characteristics can help to greatly reduce the number of instantiations as compared to the naïve approach, without loss of performance.

Firstly, in the naïve approach multiple occurrences of the same feature and operator in a single conjunction are generated, for example $A > AND\ A >$. However, this is logically identical to a rule with only one feature test, $A >$. Note that $A > AND\ A \leq$ cannot be reduced to a rule of smaller length. We can say, however, that each unique combination of feature and operator should occur only once per conjunction. For this reason, we will not use an instantiation of features and a subsequent instantiation with operators, but will combine both to a single instantiation that can occur only once per conjunction. We define a feature-operator pair (FOP) as a combination of a feature with a relational operator, e.g., $A >$.

Secondly, permutation of the FOPs in a conjunction is redundant. For example, $(A > AND\ B \leq)$ is logically equivalent to $(B \leq AND\ A >)$. Therefore, we only instantiate the FOPs in conjunctions in lexicographic order with $>$ preceding $\leq$.

Thirdly, disjunctions of conjunctions do not have to be permutated in the rules. For example, $(A > AND\ B >)\ OR\ (C > AND\ D >)$ is logically identical to $(C > AND\ D >)\ OR\ (A > AND\ B >)$. Note that $(A > AND\ B >)\ OR\ (A > AND\ B >)$ should be generated because it can be logically unique when the threshold values in the conjunctions are different. Since the conjunctions in the rule are always ordered downward in size (Alg 4.3), this problem only occurs for equally-sized conjunctions.

Note that this only holds for conjunction sizes greater than 1, since $A > \ OR\ A >$ is logically equivalent to $A >$.

Table 4.3 shows all unique instantiations of a rule of length 3 with six FOPs, taking into account the above considerations.

### Initialization of feature-operator set

The FOPs of a rule are initialized by the routine *InitFeatureOperatorSet* (Alg. 4.4) which calls *InitFeatureOperatorsConjunction* (Alg. 4.5) for each conjunction.

**Algorithm 4.4**: InitFeatureOperatorSet

```
1  for 1 ≤ i ≤ m do
2      InitFeatureOperatorsConjunction;
3  end
```

**Algorithm 4.5**: InitFeatureOperatorsConjunction

```
1  if λ_i ≠ λ_{i−1} then
       // unequally-sized conjunctions
2      for 1 < j ≤ λ_i do
3          fo_{i,j} = j;                          // lexicographically instantiate FOPs
4      end
5  else
       // equally-sized conjunctions
6      if λ_i ≠ 1 then
7          for 1 < j ≤ λ_i do
8              fo_{i,j} = fo_{i−1,j};                 // copy previous conjunction
9          end
10     else
           // size equals one
11         fo_{i,j} = fo_{i−1,j} + 1
12     end
13 end
```

Let $fo_{i,j}$ denote the index of the FOP in conjunction $i = 1 \ldots m$ at position $j = 1 \ldots \lambda_i$ in the list of all $r$ lexicographically ordered FOPs, then the ensemble of all $fo_{i,j}$ indicates a feature-operator set. As shown in Alg. 4.5, each conjunction is initialized to a lexicographically ordered set of FOPs from the list (lines 2-4). However, if the size of a conjunction is greater than 1 and equal to the size of the preceding conjunction, the conjunction is initialized by copying the instantiation of the preceding conjunction (lines 7-9). For multiple conjunctions of size 1, the conjunctions are instantiated with FOPs in lexicographical order (line 11).

**Table 4.3.** Enumeration of the instantiations of six FOPs ($A >, A \leq, B >, B \leq, C >, C \leq$) for rule length 3

| **3** | **2 OR 1** | **1 OR 1 OR 1** |
|---|---|---|
| $A > AND\ A \leq AND\ B >$ | $A > AND\ A \leq OR\ A >$ | $A > OR\ A \leq OR\ B >$ |
| $A > AND\ A \leq AND\ B \leq$ | $A > AND\ A \leq OR\ A \leq$ | $A > OR\ A \leq OR\ B \leq$ |
| $A > AND\ A \leq AND\ C >$ | $A > AND\ A \leq OR\ B >$ | $A > OR\ A \leq OR\ C >$ |
| $A > AND\ A \leq AND\ C \leq$ | $A > AND\ A \leq OR\ B \leq$ | $A > OR\ A \leq OR\ C \leq$ |
| $A > AND\ B > AND\ B \leq$ | $A > AND\ A \leq OR\ C >$ | $A > OR\ B > OR\ B \leq$ |
| $A > AND\ B > AND\ C >$ | $A > AND\ A \leq OR\ C \leq$ | $A > OR\ B > OR\ C >$ |
| $A > AND\ B > AND\ C \leq$ | $A > AND\ B > OR\ A >$ | $A > OR\ B > OR\ C \leq$ |
| $A > AND\ C > AND\ C \leq$ | $A > AND\ B > OR\ A \leq$ | $A > OR\ C > OR\ C \leq$ |
| $A \leq AND\ B > AND\ B \leq$ | $A > AND\ B > OR\ B >$ | $A \leq OR\ B > OR\ B \leq$ |
| $A \leq AND\ B > AND\ C >$ | $A > AND\ B > OR\ B \leq$ | $A \leq OR\ B > OR\ C >$ |
| $A \leq AND\ B > AND\ C \leq$ | $A > AND\ B > OR\ C >$ | $A \leq OR\ B > OR\ C \leq$ |
| $A \leq AND\ C > AND\ C \leq$ | $A > AND\ B > OR\ C \leq$ | $A \leq OR\ C > OR\ C \leq$ |
| $B > AND\ B \leq AND\ C >$ | $A > AND\ B \leq OR\ A >$ | $B > OR\ B \leq OR\ C >$ |
| $B > AND\ B \leq AND\ C \leq$ | $A > AND\ B \leq OR\ A \leq$ | $B > OR\ B \leq OR\ C \leq$ |
| $B > AND\ C > AND\ C \leq$ | $A > AND\ B \leq OR\ B >$ | $B > OR\ C > OR\ C \leq$ |
| $B \leq AND\ C > AND\ C \leq$ | $A > AND\ B \leq OR\ B \leq$ | $B \leq OR\ C > OR\ C \leq$ |
| | $A > AND\ B \leq OR\ C >$ | |
| | $A > AND\ B \leq OR\ C \leq$ | |
| | $B > AND\ B \leq OR\ A >$ | |
| | $B > AND\ B \leq OR\ A \leq$ | |
| | $B > AND\ B \leq OR\ B >$ | |
| | $B > AND\ B \leq OR\ B \leq$ | |
| | $B > AND\ B \leq OR\ C >$ | |
| | $B > AND\ B \leq OR\ C \leq$ | |
| | $B > AND\ C > OR\ A >$ | |
| | $B > AND\ C > OR\ A \leq$ | |
| | $B > AND\ C > OR\ B >$ | |
| | $B > AND\ C > OR\ B \leq$ | |
| | $B > AND\ C > OR\ C >$ | |
| | $B > AND\ C > OR\ C \leq$ | |
| | $B > AND\ C \leq OR\ A >$ | |
| | $B > AND\ C \leq OR\ A \leq$ | |
| | $B > AND\ C \leq OR\ B >$ | |
| | $B > AND\ C \leq OR\ B \leq$ | |
| | $B > AND\ C \leq OR\ C >$ | |
| | $B > AND\ C \leq OR\ C \leq$ | |
| | $C > AND\ C \leq OR\ A >$ | |
| | $C > AND\ C \leq OR\ A \leq$ | |
| | $C > AND\ C \leq OR\ B >$ | |
| | $C > AND\ C \leq OR\ B \leq$ | |
| | $C > AND\ C \leq OR\ C >$ | |
| | $C > AND\ C \leq OR\ C \leq$ | |

### Generation of next feature-operator set

We generate the next set of FOPs by increasing the index of the FOPs, $fo_{i,j}$ from right to left in the rule. We start at the last conjunction (Alg.4.6, line 1) and try to generate a new set of FOPs. If all FOP sets have been generated for this conjunction we move one conjunction to the left until we find a conjunction that can be iterated (lines 2-4). All the conjunctions to the right of this conjunction are then initialized (lines 5-10).

To generate the next FOP set for a conjunction $i$, we start with the last FOP in that conjunction and check from right to left if any FOP is the last possible for its position $j$ (Alg. 4.7, lines 1-6). For example, in Table 4.3 the last possible FOP at $j = 1$ for the conjunction of size 3 is $B \leq$, since the next FOP, C>, would not allow us to instantiate FOPs incrementally at positions $j = 2$ and $j = 3$. If $j \geq 1$ we have not arrived at the last set and we take the next FOP for this position from the ordered list and reset all FOPs to the right incrementally (lines 7-14).

### Number of feature-operator sets

Suppose we have a rule that consists of a single conjunction. In this conjunction each FOP can occur only once. The number of possible FOP sets for a single conjunction of size $\lambda$ from a set of $r$ different FOPs is equal to $\binom{r}{\lambda}$. To calculate the total number of possible FOP sets for a rule we have to take the size of each conjunction into consideration. If all conjunctions have different sizes the total number of FOP sets is the product of the number of FOP sets for the individual conjunctions. However, a rule can have conjunctions of the same size. EXPLORE generates conjunctions of equal size with replacement and without ordering from the set of $\binom{r}{\lambda}$ possible FOP combinations. This only holds for conjunctions of size $\lambda > 1$, since $A > \ OR\ A >$ is not generated. The number of FOP sets for equally-sized conjunctions, $\lambda > 1$, can be calculated by:

$$C(n, r, \lambda) = \binom{n + \binom{r}{\lambda} - 1}{n} \tag{4.5}$$

with $n$ the number of conjunctions of size $\lambda$ and $r$ the number of FOPs.

Finally, in multiple conjunctions of size $\lambda = 1$, each FOP is allowed only once. Therefore, the number of FOP sets for $n$ conjunctions with size 1 is equal to $\binom{r}{n}$.

In the case of rules that consist of different sizes of conjunctions the total number is obtained by multiplying the separate number of FOP sets, e.g. for 3 OR 2 OR 2 the total number of FOP sets is $C(1, 4, 3) \times C(2, 4, 2) = 84$.

**Algorithm 4.6**: NextFeatureOperatorSet

```
1  i = m;                                          // start at last conjunction
   // Generate Next FOPs set for the conjunction not completed yet
2  while Not (NextFeatureOperatorsConjunction) AND i ≥ 1 do
3      i = i − 1;                                  // move conjunction to the left
4  end
5  if i ≥ 1 then
       // initialize all conjunctions to the right
6      while i < m do
7          i = i + 1;
8          InitFeatureOperatorsConjunction;
9      end
10     Return TRUE;
11 else
12     Return FALSE;                               // all FOP sets have been generated
13 end
```

**Algorithm 4.7**: NextFeatureOperatorsConjunction

```
1  j = λ_i ;                                       // start at last position in conjunction
   // determine which j is not at last possible FOP index
2  last = r;
3  while fo_{i,j} = last do
4      j = j − 1;
5      last = last − 1;
6  end
   // insert FOPs from current position (j ≥ 1) till last position
7  if j ≥ 1 then
8      fo_{i,j} = fo_{i,j} + 1;
9      while j < λ_i do
10         j = j + 1;
11         fo_{i,j} = fo_{i,j−1} + 1;
12     end
13     Return TRUE;                                // next set of FOPs generated
14 else
15     Return FALSE;                               // last set of FOPs generated
16 end
```

Up to now we only calculated the number of FOP sets for a specific logical-operator set. The total number of FOP sets for a certain rule length can be calculated by repeating the calculation described above for each logical-operator set. There is a large reduction compared to the naïve exhaustive search (Equation 4.3). For example, Fig. 4.1 shows the ratio of the number of rules instantiated with FOPs in the naïve approach and the number of rules in the approach used in EXPLORE for 6, 10, and 14 FOPs, up to rule length 5.

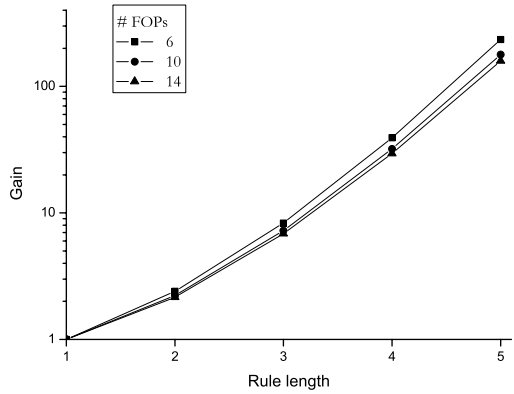


**Figure 4.1.** Reduction of number of feature-operator instantiations in the new algorithm compared to the naïve approach for 6, 10, and 14 FOPs at different rule lengths.

As shown in Fig. 4.1 the gain increases with the rule length because the naïve approach generates more unnecessary repetitions with increasing rule length. Furthermore, for a certain rule length the gain is higher if the number of FOPs is lower. This is the result of the unnecessary repetitions of identical conjunctions in the naïve approach, which are skipped in the EXPLORE algorithm, e.g., for logical-operator set 1 $OR$ 1 $OR$ 1 the gain equals $r^3/\binom{r}{3}$ which is higher for lower values of $r$.

### 4.2.3 Generation of threshold sets

The next task is to instantiate the threshold values in the rules that up to now consist of logical operators and FOPs. Because the number of potential threshold values is often very large compared to the number of FOPs, especially for continuous-valued features, the total number of rules that are generated is mainly determined by the number of potential threshold values per feature. Next we propose two methods to greatly reduce the search space at the threshold instantiation level of EXPLORE.

**Determination of candidate threshold values**

Since EXPLORE performs an exhaustive search, the simplest method would be to use all measurements in the data set as potential threshold values. The question is, whether we need to look at all these values to perform an exhaustive search. The answer is that a large reduction in threshold values is possible if the principals of subsumption pruning, often used in beam search strategies, are applied to the discretization problem.

In subsumption pruning a rule is removed from the set of promising rules if there exists another rule that covers a superset of the cases that belong to the class for which the rule is developed (positive cases) and a subset of cases that belong to the other class (negative cases) [9]. For example, if we have the following two rules:

1. if $A > 5$ then $class = 1$
2. if $B > 7$ then $class = 1$

If rule 1 covers a superset of the positive cases covered by rule 2, and a subset of the negative cases, then rule 1 has a higher coverage of correct cases. Any conjunctive extension of rule 1 with a second feature test will be better than the same extension of rule 2.

The ideas behind subsumption pruning can also be used for our purpose to reduce the number of potential threshold values without degradation of performance. Suppose we have a small data set ordered by increasing feature values of a feature A as shown in Table 4.4, and we want to determine the candidate threshold values.

**Table 4.4.** Small data set consisting of 6 cases with values of a feature A and their class label.

| Feature A | Class |
|---|---|
| 4.50 | 0 |
| 5.00 | 0 |
| 5.10 | 1 |
| 5.20 | 1 |
| 5.40 | 0 |
| 6.00 | 1 |

The simplest solution would be to take all 6 feature values as candidate threshold values for both FOPs ($A >$ and $A \leq$). However, it does not pay off to select threshold

values at adjacent feature values that assign to the same class, because the separation of the classes will never improve at those values. Meaningful threshold values can only occur between the feature values of two adjacent, but differently labeled cases. This has previously been shown by Fayyad for greedy approaches [15]. However, there is a larger reduction possible if we use the subsumption principle as well. To illustrate this we calculate for FOP $A >$, the number of correct positives $(CP)$ and false positives $(FP)$ for all average feature values of adjacent cases as shown in Table 4.5.

**Table 4.5.** The number of correct positives (CP) and false positives (FP) taking the average feature values of adjacent cases as the threshold values for the $>$ operator.

| FOP A > | CP | FP |
|---|---|---|
| >4.75 | 3 | 2 |
| >5.05 | 3 | 1 |
| >5.15 | 2 | 1 |
| >5.30 | 1 | 1 |
| >5.70 | 1 | 0 |

Based on the subsumption principle, only those threshold values should be used that cover a superset of the positive cases and a subset of the negative cases. Thus, in our example, $> 5.05$ is preferred above $> 4.75$, because there are less $FPs$ for the same number of $CPs$. This also holds for $> 5.15$, since the same number of $FPs$ are covered but less $CPs$. In our example the final candidate threshold values are 5.05 and 5.70. For the relational operator $>$, only threshold values that are at a class boundary from 0 to 1 need to be taken as candidate threshold values for EXPLORE. Similarly, for the relational operator $\leq$ we only need to include threshold values that are at a class boundary from 1 to 0. However, if there are cases in the data set with equal feature values but with different class labels, these boundaries have to be taken as candidate threshold values for both relational operators.

Application of the subsumption principle thus allows us to remove a large number of measurement values from the list of candidate threshold values for each FOP. In Table 4.6, the effect of different threshold selection strategies on the total number of threshold values of all FOPs are presented for 5 well-known data sets taken from the UCI data repository [16].

**Table 4.6.** Number of candidate threshold values for five standard data sets using three different approaches of threshold value generation: exhaustive, class boundaries, and the subsumption principle. The percentage reduction of the class boundaries and subsumption approach compared to exhaustive generation is indicated in parentheses.

| | | #Threshold values | | |
|---|---|---|---|---|
| Data set | #Features | Exhaustive | Boundaries | Subsumption |
| Heart | 13 | 756 | 580 (-23.3%) | 426 (-43.7%) |
| Hepatitis | 19 | 542 | 232 (-57.2%) | 152 (-72.0%) |
| Diabetes | 8 | 2496 | 1714 (-31.3%) | 1174 (-53.0%) |
| Breast | 9 | 178 | 144 (-19.1%) | 130 (-27.0%) |
| Liver | 6 | 654 | 538 (-17.7%) | 432 (-34.0%) |

**Threshold instantiation by branch and bound**

After the candidate threshold values have been selected EXPLORE will instantiate the rules with the threshold values in a systematic manner using a branch and bound approach. Branch and bound is a technique that is often used to solve large-scale combinatorial optimization problems [9]. The complete space of solutions is iteratively searched by applying branching and bounding rules. The first operation of an iteration is branching, i.e., the unexplored solution space is subdivided into two or more subspaces to be investigated in a next iteration. Subsequently, a bounding function for each of the subspaces is calculated and compared to the current best solution. If it can be established that a subspace cannot contain the optimal solution, this subspace is discarded, else it is further explored and the solution replaces the current best. The search terminates when there is no unexplored part of the solution space left, and the optimal solution is the then current best solution. The branch and bound technique can provide a tightly focused traversal of the search space, while assuring that the optimal solution will be found.

In our threshold instantiation problem we need to systematically instantiate all combinations of threshold values for all FOPs in each rule. To apply the branch and bound technique to this problem we need to define bounding rules. Clearly, the bounds are dependent on the performance measure to be optimized and on the user-defined performance constraints. We will present the definition of the bounding rules for two

optimization problems: sensitivity optimization with a constraint on specificity, and accuracy optimization.

**Sensitivity optimization with constraints on specificity**

In sensitivity optimization the number of CPs needs to be maximized. We define the feature test $ft_{i,j}$ as the combination of the FOP and the threshold value of conjunction $i$ at position $j$. An important consideration is that the maximum number of cases that are correctly assigned to the positive class by conjunction $i$ ($CPmax_i$) is determined by the feature test that has the lowest number of CPs ($CP_{i,j}$).

$$CPmax_i = min(CP_{i,1}, \ldots, CP_{i,\lambda_i}) \tag{4.6}$$

Suppose we have a rule that consists of one conjunction ($A > AND\ B >$). If a feature test $A > 5$ correctly classifies 10 cases as belonging to the positive class ($CP_{1,1} = 10$) and the current best rule classifies 15 cases correctly ($CPbest = 15$), then no instantiation of threshold values in the subsequent features test ($B >$) will ever result in a decision rule with a higher sensitivity than the current best. All conjunctive extensions of $A > 5$ can be skipped without loss of performance. Note that even if all feature tests have a $CP_{i,j} > CPbest$ we cannot conclude that the whole conjunction will have a better $CP$ than the current $CPbest$, because the same cases can be classified correctly by one feature test and incorrectly by another.

Now suppose we have a rule that consists of more than one conjunction. Then we can use a similar approach for each conjunction but the minimum number of CPs of conjunction $i$ ($CPmin_i$), depends on $CPbest$ and the number of CPs of the other conjunctions. Let $CP_k$ be the cumulative number of correct positives of the rule up to conjunction $k$. For example, if we have a logical-operator set 2 $OR$ 2 $OR$ 2, $CP_2$ equals the number of $CPs$ of the disjunction of the first two conjunctions. In EXPLORE we modify threshold values in the rule starting at the last FOP in the last conjunction while the other FOPs are not changed and we move to the left if this is necessary based on the bounding rules. Therefore, each cumulative $CP_i$ can be calculated once and used as long as no feature tests are changed in that particular ensemble of conjunctions. We can now calculate the $CPmin_i$ bound by taking into account the cumulative number of CPs up to conjunction $i-1$, $CP_{i-1}$, as well as the $CPmax_i$ of the conjunctions $i+1, \ldots, m$:

$$CPmin_i = CPbest - CP_{i-1} - \sum_{k=i+1}^{k=m} CPmax_k \tag{4.7}$$

The next step is to also define a bound based on the specificity constraint and reduce the search space even further. By defining a constraint on specificity the user actually allows a maximum number of cases that may falsely be classified as positive ($FPmax$). We can use this as a bound on the number of FPs of a conjunction $i$, $FP_i$, i.e., $FP_i \leq FPmax$. Note that we now cannot use a different $FPmax$ for each conjunction since the same cases can be classified as false positives by multiple conjunctions. This $FPmax$ bound is a constant value defined by the user, in contrast to the $CPmin$ bounds which are updated after a new threshold set is generated and the performance is calculated.

### Initialization of threshold values

The first threshold set is generated by the algorithm *InitThresholdSet* (Alg. 4.8) which calls *InitThresholdsConjunction* (Alg. 4.9) for each conjunction.

**Algorithm 4.8**: InitThresholdSet

```
1  for 1 ≤ i ≤ m do
2      InitializeThresholdsConjunction;
3  end
```

**Algorithm 4.9**: InitThresholdsConjunction

```
1  for 1 ≤ j ≤ λ_i do
2      th_{i,j} = 1 ;                  // Initialize to the first threshold in the list
3  end
   // Check CP and FP Bound
4  if CP_{i,j} < CPmin_i then
5      Return FALSE ;                  // Conjunction can never fulfill CP bound
6  end
7  if FP_i > FPmax then
8      if NextThresholdsConjunction then
9          Return TRUE;
10     else
11         Return FALSE
12     end
13 else
14     Return TRUE;
15 end
```

The *InitThresholdsConjunction* routine (Alg. 4.9) initializes to the first set of threshold values that fulfills the current bounds. Let $th_{i,j}$ be the index in the list of candidate threshold value and $nt_{i,j}$ the number of candidate threshold values of FOP $fo_{i,j}$. For

each threshold value of all feature-operators the number of CPs are calculated beforehand and ordered downwards. Without prior knowledge, the threshold values with the highest number of CPs appear to be the best threshold set to start with since it has the highest potential to improve on *CPbest* (lines 1-3). If the CP bound is not met we return "FALSE" because a higher number of CPs will not be possible (line 4-6). Then we check the FP bound and, if $FP_i > FPmax$, call the *NextThresholdsConjunction* routine (Alg. 4.11), which generates the next set of threshold values that fulfills both bounds (line 7-13).

### Generation of next threshold set

The *NextThresholdSet* routine instantiates the next set of threshold values for the complete rule (Alg. 4.10). It starts with the last conjunction and generates the next set of threshold values for that conjunction, if possible (line 4). If this is not possible we move to the left until a conjunction is found that can be instantiated with a new set of threshold values (line 11). Each time a conjunction $i$ can be instantiated we have to initialize all conjunctions to its right again (lines 6-9). However, it is possible that one of these conjunctions cannot be initialized to a value that fulfills the bounds. In that case, we determine the next threshold set for conjunction $i$ and try again. The routine will return "TRUE" if a rule is found that fulfills both bounds.

**Algorithm 4.10**: NextThresholdSet

```
1  i = m;                                          // Start with last conjunction
2  result=FALSE;
3  while i > 1 AND NOT result do
4      if NextThresholdsConjunction then
5          result=TRUE;
           // initialize conjunctions i+1,...,i = m
6          while i < m and result do
7              i = i + 1;
8              result=InitThresholdsConjunction;
9          end
10     else
11         i = i − 1;
12     end
13 end
14 Return result;
```

To generate the next threshold set for a conjunction (Alg. 4.11), we start with the last FOP (line 1) and search for the FOP for which there still exists a next threshold value

($th_{i,j} < nt_{i,j}$) and has a $CP_{i,j} \geq CP_{min}$ (line 2-5). We then increment the threshold index for that FOP and evaluate the complete conjunction. If the bounds are not met we call the routine again.

**Algorithm 4.11**: NextThresholdsConjunction

```
1  j = λ_i;                                                    // start with last FOP
   // Find FOP that is not at its last threshold value and CP_{i,j} ≥ CPmin_i
2  while (th_{i,j} = nt_{i,j} OR CP_{i,j} < CPmin_i) AND j ≥ 1 do
3      th_{i,j} = 1;                                           // reset to first threshold
4      j = j − 1;                                              // move one FOP to the left
5  end
6  if j ≥ 1 then
7      th_{i,j} = th_{i,j} + 1;                                // instantiate next threshold value
8      Evaluate Conjunction;                                   // determine CPs and FPs of the conjunction
9      if CP_i < CPmin_i OR FP_i > FPmax then
10         NextThresholdsConjunction;
11     else
12         Return TRUE;
13     end
14 end
15 Return False;
```

## Execution time with and without branch and bound

In Table 4.7 the effect of the branch and bound approach on the execution time is illustrated if we generate rules up to length $n = 3$. The sensitivity of the rule is optimized with minimum specificities of 90%, 95%, and 99%. The execution time is compared to the procedure without branch and bound optimization. EXPLORE was implemented on an Intel Pentium 1.7 GHz with 1.0 Gb system memory.

From Table 4.7 it is apparent that the execution time is greatly reduced by the branch and bound approach, i.e. a very large number of combinations of threshold values is skipped without loss of performance. Execution time increases with increasing specificity bound, which at first might seem counter-intuitive. However, the higher constraint on specificity will lower the *CPbest* and thus reduces the effect of the *CPmin* bounds within all conjunctions.

**Table 4.7.** Execution time (hour:min:sec) of EXPLORE without the use of branch and bound (BB), and with BB when rules up to rule length 3 are optimized for sensitivity at different specificity constraints of 90%,95%, and 99% (Sp90, Sp95, and Sp99, respectively).

| | | | | with BB | | |
|---|---|---|---|---|---|---|
| Data set | #Features | #Cases | without BB | Sp90 | Sp95 | Sp99 |
| Heart | 13 | 270 | 0:06:06 | 0:00:08 | 0:00:21 | 0:00:22 |
| Hepatitis | 19 | 155 | 0:00:20 | 0:00:02 | 0:00:03 | 0:00:03 |
| Diabetes | 8 | 768 | 2:37:56 | 0:03:27 | 0:06:53 | 0:14:30 |
| Breast | 9 | 699 | 0:00:17 | 0:00:02 | 0:00:02 | 0:00:02 |
| Liver | 6 | 345 | 0:06:23 | 0:00:22 | 0:00:28 | 0:00:41 |

**Accuracy optimization**

It is also possible to use the branch and bound principle for the optimization of the accuracy. Accuracy is determined as the total number of correct classifications, $CT = CP + CN$. If we assume that all negative cases are correctly classified then we can calculate for each conjunction the minimal number of CPs, $CPmin_i$, to improve on the current number of correctly classified cases, $CTbest$. $CPmin_i$ for each conjunction can be calculated by taking into account the cumulative number of CPs up to conjunction $i-1$, $CP_{i-1}$, as well as the $CPmax_i$ of the conjunctions $i+1, \ldots, m$:

$$CPmin_i = CTbest - \#Negative\ cases - CP_{i-1} - \sum_{i+1}^{i=m} CPmax_i \tag{4.8}$$

An FPmax bound can be defined if we assume that all positive cases are correctly classified:

$$FPmax = \#cases - CTbest \tag{4.9}$$

We can now use the same algorithm as described for the sensitivity optimization with specificity constraint but with different values for the CPmin's and for FPmax. These bounds are less strong than for sensitivity optimization because they are defined for

the extreme case that all negative or positive cases are classified correctly. Nevertheless, the reduction in execution time is still large as shown in Table 4.8.

**Table 4.8.** Execution time (hour:min:sec) of EXPLORE without the use of branch and bound (BB), and with BB when rules up to rule length 3 are optimized for accuracy.

| Data set | #Features | #Cases | without BB | with BB |
|---|---|---|---|---|
| Heart | 13 | 270 | 0:06:06 | 0:00:41 |
| Hepatitis | 19 | 155 | 0:00:20 | 0:00:05 |
| Diabetes | 8 | 768 | 2:37:56 | 0:29:32 |
| Breast | 9 | 699 | 0:00:17 | 0:00:02 |
| Liver | 6 | 345 | 0:06:23 | 0:02:39 |

## 4.3 Experiments

We compared the performance of EXPLORE up to rule length $n = 3$ with two decision-tree algorithms, C4.5 [5] and CART [4]. Table 4.9 shows the average accuracy and standard error of ten runs of a 10-fold cross-validation experiment on the data sets used before. Before each run we first randomized each of the data sets into 10 folds and used the same folds for all algorithms.

Most exhaustive search results are comparable to the best classifiers generated by the other two algorithms, but for the Heart data set a higher performance is obtained with EXPLORE. By using the same folds we were able to compare the classifiers generated by the different algorithms for each fold. For the Diabetes and Liver data set, the rules generated by EXPLORE often consisted of the same FOP sets with slightly different threshold values. The classifiers generated by CART and C4.5 vary more in size of, over the folds. Generally, these algorithms show larger variability in accuracy than EXPLORE, as expressed by lower standard errors for EXPLORE.

Since EXPLORE is the only of the three algorithms that can optimize sensitivity under specificity constraints, it is not possible to compare the algorithms in this respect. As an illustration, Fig. 4.2 shows the Receiver Operator Curve (ROC) for the Heart data set as generated by EXPLORE for rule lengths up to $n = 3$ if we use all available

**Table 4.9.** Comparison of average accuracy (standard error) of the three induction algorithms on five standard data sets.

| | | | EXPLORE | | |
|---|---|---|---|---|---|
| Data set | CART | C4.5 | n=1 | n=2 | n=3 |
| Heart | 78.8 (0.5) | 78.5 (0.5) | 73.9 (0.4) | 73.5 (0.6) | 82.1 (0.2) |
| Hepatitis | 87.7 (0.6) | 82.6 (0.8) | 80.4 (0.5) | 86.9 (0.7) | 87.5 (0.7) |
| Diabetes | 75.9 (0.2) | 74.2 (0.5) | 73.7 (0.1) | 76.7 (0.1) | 75.5 (0.2) |
| Breast | 95.2 (0.2) | 95.6 (0.2) | 91.6 (0.2) | 95.3 (0.1) | 94.9 (0.1) |
| Liver | 68.9 (0.5) | 64.9 (0.7) | 61.2 (0.4) | 68.3 (0.3) | 66.4 (0.3) |

data for learning. This curve was generated by repeatedly optimizing sensitivity with EXPLORE with an increasing constraint on specificity in steps of 5%.

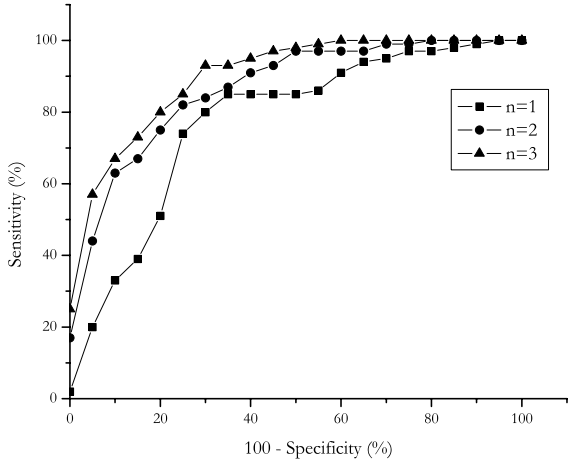


**Figure 4.2.** Receiver operator curves for the Heart data set up to rule length $n = 3$.

## 4.4 Discussion

We developed an algorithm, EXPLORE, that performs an exhaustive search to find a decision rule that optimizes a performance measure while enforcing user-specified constraints on other performance measures. The execution time of the algorithm could be significantly reduced compared to a more naïve approach by incorporating several new techniques that greatly reduce the search space. Firstly, the instantiation of feature-operator pairs instead of features and operators separately enables to generate rules in a more efficient and systematic manner. Secondly, application of the subsumption principle reduced the number of candidate thresholds in the five standard data sets by a third on average as compared to only taking thresholds at the class boundaries. Thirdly, the use of the branch and bound techniques assures that no rules are generated and evaluated that cannot have a higher performance than the currently best performing rule. Note that the execution time of the naïve search would greatly exceed those presented in Table 4.7 for EXPLORE without the use of branch and bound since in the latter algorithm we still make use of the reduction in thresholds and the FOP approach.

On five standard data sets from the medical domain the accuracy of the classifiers generated by EXPLORE are comparable to or surpasses the accuracy of the classifiers generated by the greedy algorithms C4.5 and CART. More importantly, however, is the capability of EXPLORE to optimize diagnostic performance measures under constraints. In CART it is possible to steer the trade-off between sensitivity and specificity with the use of cost functions [4]. In practice this is often a trial-and-error approach, because the relation between the cost functions and their effect on diagnostic performance is unknown. To our knowledge, the only algorithm that allows control over more than one performance measure is the PVM algorithm [7], but PVM performs a heuristic beam search strategy, not an exhaustive search. The control over other performance measures is important, especially in medical problems where a balance must be struck between sensitivity and specificity. In the medical data sets that we used to validate EXPLORE, the highest accuracy value often leads to a position on the ROC curve that has no value in medical practice. With EXPLORE we can enhance the sensitivity of the rule by lowering its specificity while imposing minimum constraints on the accuracy. However, the performance measures are not independent meaning that the various minimum constraints should be selected carefully, i.e. when more than one constraint is imposed it is possible that no solution exists.

There are several ways to further increase the induction speed of EXPLORE. Currently we are developing a distributed version of EXPLORE that runs in parallel on a grid of computers. The systematic rule generation of EXPLORE enables parallel computa-

tion on different levels of the algorithm. Another approach is to use expert knowledge in the induction process. In the version of EXPLORE described here we did not incorporate any available knowledge to guide the search. However, EXPLORE lends itself very well to the incorporation of expertise, which can lead to a large reduction in complexity without loss of performance. For example, based on prior knowledge some features can be made obligatory, ranges for candidate threshold values can be specified, or rules can be defined as a starting point for induction. The expert may also predefine a minimum value of the sensitivity by increasing the CPmin bound beforehand and then optimize sensitivity with a constraint on the specificity. This can improve the speed of the search considerably because a higher bound at the start of the induction reduces the number of potential rules. Another advantage of the use of expert knowledge is that this may also improve the comprehensibility of the resulting rule. Comprehensibility is generally greater for smaller rules and for rules that contain features that are easily interpretable for the end-user. Finally, heuristics can be applied to improve the speed. For example, the number of reoccurences of a FOP in a rule could be restricted. Our future research, will focus on the enhancement of the processing speed of the algorithm.

## 4.5 References

[1] Briscoe G, Caelli T. A compendium of machine learning volume 1: symbolic machine learning. Norwood, New Jersey: Ablex Publishing Corporation, 1996.

[2] Mitchell T. Machine learning. McGraw Hill, 1997.

[3] Clark P, Niblett T. The CN2 induction algorithm. Machine Learning 1989;3:261–283.

[4] Breiman L, Friedman JH. Classification and regression trees. Belmont, CA: Wadsworth and Brooks, 1984.

[5] Quinlan JR. C4.5: programs for machine learning. San Mateo, CA: Morgan Kaufmann, 1992.

[6] Galen R, Gambino S. Beyond normality: the predictive value and efficiency of medical diagnosis. New York: Wiley, 1975.

[7] Weiss S, Galen R, Tadepalli P. Maximizing the predictive value of production rules. Artificial Intelligence 1990;45:47–71.

[8] Segal RB. Machine learning as massive search. Ph.D. thesis, University of Washington, 1997.

[9] Webb GI. OPUS: an efficient admissible algorithm for unordered search. Journal of Artifical Intelligence 1995;3:431–465.

[10] Holte R. Very simple classification rules perform well on most commonly used datasets. Machine Learning 1983;11:63–90.

[11] Kors JA, Hoffmann AL. Induction of decision rules that fulfil user-specified performance requirements. Pattern Recognition Letters 1997;19:1187–1195.

[12] Andrews GE. The theory of partitions. In: Rota G, editor. Encyclopedia of mathematics and its applications. Reading, MA: Addison-Wesley, 1976; pp. 214–216.

[13] Andrews GE, Eriksson K. Integer partitions. New York: Cambridge University Press, 2004.

[14] Zoghbi A, Stojmenovic I. Fast algorithms for generating integer partitions. Intern J Computer Math 1998;70:319–332.

[15] Fayyad U, Irani K. On the handling of continuous-valued attributes in decision tree generation. Machine Learning 1992;8:87–102.

[16] Murphy P, Aha D. UCI repository of machine learning databases. Department of Information and Computer Science. University of California, Irvine (http://www.ics.uci.edu/ mlearn/MLRepository.html), 1994.

# 5 PEDMEANS: a computer program for the interpretation of pediatric ECGs

P.R. Rijnbeek*, M. Witsenburg†, A. Szatmari‡, J. Hess§, J.A. Kors*

* Dept. of Medical Informatics, Faculty of Medicine and Health Sciences, Erasmus University, Rotterdam, The Netherlands
† Dept. of Pediatric Cardiology, Sophia Children's Hospital, Rotterdam, The Netherlands
‡ Dept. of Pediatric Cardiology, Hungarian Institute of Cardiology, Budapest, Hungary
§ Dept. of Pediatric Cardiology and Congenital Heart Disease, Deutsches Herzzentrum München, Munich, Germany

Journal of Electrocardiology Vol. 34 Supplement 2001

## Abstract

The interpretation pediatric electrocardiograms (ECGs) is complicated because of the strong age-dependency of the diagnostic criteria. We wanted to develop and evaluate a computer program for the interpretation of pediatric 12-lead ECGs.

Continuous age-dependent normal limits were established based on ECGs from 1912 healthy Dutch children. Additionally, a reference interpretation was obtained for 1718 ECGs recorded at the Sophia Children's Hospital. The total set of ECGs was divided in a training set of 1076 ECGs and a test set of 642 ECGs. All ECGs were recorded at a sampling rate of 1200 Hz. Based on the normal limits and the training set, diagnostic rules were formalized in an iterative process, using expert interviews and automatic rule induction. The resultant rules were evaluated on the test set.

The performance of the program, on our study population, appears to justify its use in a clinical setting. Preferably, the program should also be evaluated in other clinical centers.


**Keywords:** computer program, electrocardiography, pediatrics

## 5.1 Introduction

While much research has been devoted to computerized interpretation of the adult electrocardiogram (ECG), automatic interpretation of pediatric ECGs has received little attention. It has been argued, however, that computer support in pediatric electrocardiography would be even more beneficial because the interpretation is more complex due to the strong age-dependency of the diagnostic criteria [1–3]. Especially in the first year of life the electrical behaviour of the heart changes rapidly [4,5]. This is mainly the result of the change from right to left ventricular predominance, which is reflected in changes in normal wave amplitudes. In addition, intervals are markedly reduced in the young child and gradually prolong to adult values as the child grows. This age-dependency forces the pediatric cardiologist to make use of extensive tables of normal values. A computer program could be helpful in relieving the pediatric cardiologist from memorizing the large number of age-dependent criteria. Also, a computer program would make possible interpretation of pediatric ECGs at places currently devoid of cardiologic experience.

A small number of computer programs have been developed for the interpretation of the pediatric ECG [2,6–9] or vectorcardiogram [1,3,10–12]. In some early systems [6–8], leads were recorded and processed in groups of three simultaneously, and the ECG signals were sampled at 250 Hz. Nowadays, modern electrocardiographs record 12 simultaneous leads at higher sampling rates that have been shown to avoid significant errors, especially in wave amplitudes [5,13]. More recently, Macfarlane extended the Glasgow program with a pediatric ECG interpretation module for simultaneous 12-lead ECGs, sampled at a rate of 500 Hz [9]. Unfortunately, the performance of this program has not yet been reported. Also, it has been questioned whether even 500 Hz is a high enough sampling rate for the recording of pediatric ECGs [14–16].

For the development of a computer program for pediatric ECG interpretation, up-to-date normal limits are essential. Several studies have been conducted to determine normal limits for the pediatric ECG [4,13,17–21]. However, all these studies have certain imperfections that limit their practical applicability. Firstly, normal limits have often been presented for an incomplete set of clinically relevant parameters and leads. Secondly, in many studies parameters were measured by hand from ECGs recorded on paper. At present, computer analysis of digitised ECGs allows more accurate measurement. Thirdly, in some studies the ECG signals may have been recorded less than perfectly owing to low sampling rates or the use of ECG amplifiers with small bandwidth. Therefore, the establishment of new up-to-date normal limits would seem necessary.

This study describes the development and evaluation of a computer program for the interpretation of pediatric ECGs, based on a large set of normal and abnormal ECGs recorded at a high sampling rate of 1200 Hz.

## 5.2 Materials and Methods

Our newly developed pediatric ECG interpretation program, which is called PEDiatric Modular ECG ANalysis System (PEDMEANS), is based on the Modular ECG ANalysis System (MEANS) for the interpretation of adult ECGs [22]. We had to modify both the signal analysis and the classification sections of MEANS to allow for pediatric ECG interpretation. The signal analysis section takes care of ECG-complex detection and typification, selective averaging, waveform recognition, and measurement of ECG parameters, while the classification section deals with rhythm analysis and morphological (contour) interpretation. Children, as compared to adults, have higher heart rates, narrower QRS complexes, and the ECGs are often noisier and show more artifacts. To take care of these differences, we modified several digital filters and detection algorithms. Also, changes were made in the measurement part of MEANS. Furthermore, since MEANS has been developed for ECGs sampled at 500 Hz, the signal analysis algorithms were changed to comply with the higher sampling rate of 1200 Hz. The rhythm analysis of MEANS was basically left the same, but threshold values were adjusted and made age-dependent. Because of the large differences in contour interpretation between adults and children, this part has completely been redeveloped for PEDMEANS. To that end, we established new normal limits for pediatric ECGs and developed decision rules for all diagnostic categories.

#### Continuous age-dependent normal limits

Data collection and analysis for the establishment of normal limits have been described previously [5]. Briefly, 1912 children aged 11 days to 16 years were recruited at three child health centers, three primary schools, and a secondary school in Rotterdam. Children with previously known cardiovascular abnormalities were excluded from the study. For each child, a 12-lead ECG was recorded using a portable PC-based acquisition system (Cardio Control, Delft, the Netherlands) at a sampling rate of 1200 Hz and then processed by PEDMEANS. To reduce noise, PEDMEANS computes a representative averaged beat for each of the twelve leads, from which ECG measurements are derived. The 2nd and 98th percentiles of the measurement distribution

were taken as the lower limit and the upper limit of normal, respectively. In addition to the estimation of normal limits for age groups, we determined age-dependent curves that represent the normal limits in a continuous form, to avoid abrupt changes in diagnosis with small differences in age. In a window of 200 measurements, moving along the age axis, percentiles and their confidence intervals were calculated for each window position according to a two-stage transformation procedure [5]. Polynomial curves were then fitted through the 2nd and 98th percentile values to obtain percentiles that smoothly change with age. The order of the polynomials was determined by visual inspection of the fit, selecting the lowest order that yielded curves remaining within the estimated confidence intervals. As an example, Fig. 5.1 shows the estimated continuous age-dependent normal curves for QRS duration.

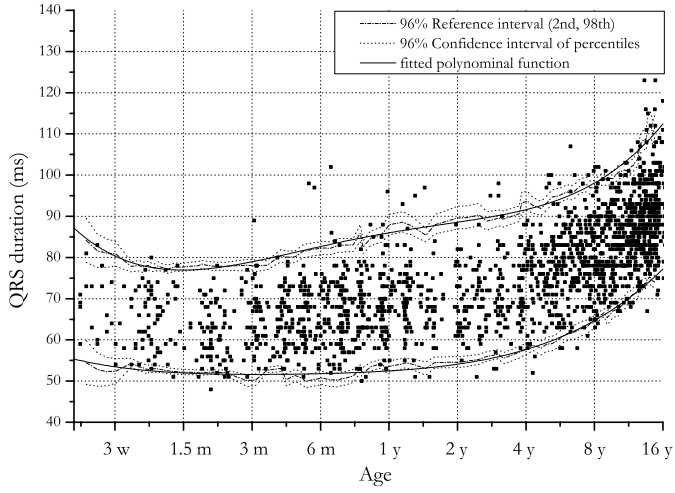


**Figure 5.1.** Continuous age-dependent percentile curves of the QRS duration.

### Database

To develop and validate the diagnostic rules, 1718 ECGs were recorded at the Sophia Children's Hospital in Rotterdam with the Cardio Control equipment described above. Two pediatric cardiologists independently interpreted all ECGs and rated the certainty of each abnormality on a four-point scale: absent (=0), possible (=1), probable (=2), and definite (=3). If the interpretations of the two pediatric cardiologists differed by only one point, either of them was randomly selected as the final reference interpretation. If the interpretations differed by two or three qualifier points, a third pediatric

cardiologist adjudicated the ECG and his interpretation was then taken as the reference. The ECGs were divided in a training set of 1076 ECGs and a test set of 642 ECGs. In Table 5.1, the number of cases in the categories left ventricular hypertrophy (LVH), right ventricular hypertrophy (RVH), left bundle branch block (LBBB), and right bundle branch block (RBBB) are shown. The remaining cases are a mixture of different other abnormalities and normals. Based on the training set, diagnostic rules were developed through expert interviews and with the use of a modified version of our EXPLORE (Exhaustive Procedure for Logic-Rule Extraction) induction algorithm [23].

**Table 5.1.** Number of cases in the categories: right ventricular hypertrophy (RVH), left ventricular hypertrophy (LVH), right bundle branch block (RBBB), and left bundle branch block (LBBB), in the training set and test set.

| | Training set | | | Test set | | |
|---|---|---|---|---|---|---|
| Abnormality | Possible | Probable | Definite | Possible | Probable | Definite |
| RVH | 50 | 29 | 52 | 76 | 40 | 34 |
| LVH | 51 | 32 | 33 | 80 | 21 | 13 |
| RBBB | 5 | 21 | 67 | 11 | 18 | 32 |
| LBBB | 0 | 2 | 6 | 0 | 0 | 4 |

### Automatic Rule induction

The elucidation of expert knowledge for decision-support systems is a tedious and time-consuming task. Whereas experts are generally well able to classify a set of given objects, they often find it difficult to articulate the knowledge they are using to do this, the more so if that knowledge has to be precisely formulated in order to be implemented in a computer program. To alleviate this problem, induction techniques can be used that automatically construct a classifier based on a set of labeled objects [24]. Our EXPLORE (Exhaustive Procedure for LOgic-Rule Extraction) induction algorithm generates decision rules that fulfill user-specified performance constraints. Most induction algorithms try to maximize the accuracy of the classifier that is being derived, whereas in medicine a classifier often needs to be optimal with respect to other diagnostic performance measures, such as sensitivity or specificity. Basically, EXPLORE generates all possible rules and searches for the one with the best performance. To reduce the number of logical combinations to be generated, several heuristics can be employed that reduce the number of features, thresholds, or relational operators [23]. A further reduction of search complexity was obtained by adding several options that take advantage of prior expert knowledge. This may also improve the comprehensibility of the resulting rule. First, we modified EXPLORE to allow the expert to indicate

which features must be present in a decision rule for the rule to make sense. For example, in RVH the R-wave amplitude in $V_1$ should definitely be present in the decision rule. Second, the expert may define ranges of threshold values and let the induction algorithm find the optimum threshold within the specified range. Such a specification may for example be based on reasonable assumptions about normal limits of parameters. Third, the expert may specify subrules that partially describe an abnormality and are a starting point for further induction by EXPLORE. For example, in right ventricular hypertrophy the R-wave amplitude in lead $V_1$ and the S-wave amplitude in $V_6$ are known to be important, but a pediatric cardiologist may find it difficult to give precise threshold values. In the extended version of EXPLORE, the following subrule could then be defined:

$$(R\ wave\ \mathrm{V}_1 > ?\ AND\ S\ wave\ \mathrm{V}_6 > ?) \tag{5.1}$$

where the question marks indicate threshold values to be derived by EXPLORE. The algorithm will then only consider rules that contain this subrule. The expert may also replace operators by question marks, or may specify complete subrules without question marks, in concordance with his prior knowledge.

**Rule development**

Using the learning set, we tried to maximize the sensitivity of the decision rules while keeping specificity at $\geq 95\%$, a level which we considered necessary for the rules to be of practical utility. This was accomplished by an iterative procedure in which a pediatric cardiologist was asked to articulate his knowledge in the EXPLORE formalism, EXPLORE generated the best decision rule, and the comprehensibility and performance of this rule was discussed with the expert, possibly resulting in refinement of the specified prior knowledge. These steps could be reiterated until a satisfactory rule was obtained.

**Performance evaluation**

The test set was used to assess the performance of the final rules. Accuracy, sensitivity, and specificity for each diagnostic category were calculated from 2-by-2 classification matrices. Category dichotomization was performed by mapping qualifiers "probable" and "definite" to "present", and "possible" and "absent" to "not present".

## 5.3 Results

To illustrate our rule-development approach, we will elaborate on the decision rule for RBBB. In RBBB, the right ventricle is not depolarized directly through the Purkinje system, but through the ventricular myocardium at a much slower rate due to the slower conduction velocity in the myocardium. This results in a sequential depolarization of the left and right ventricles and a terminal depolarisation that is directed rightward and anteriorly. The ECG shows a QRS duration that is longer than normal for age, wide and slurred R′ waves in $V_{3R}$, $V_1$, and aVR, and wide and slurred S waves in I, II, $V_6$, and $V_7$. After studying all the RBBB cases in the training set during the expert interviews, it appeared that the pediatric cardiologists looked at the total QRS duration, the ratio of R′ and S durations in $V_{3R}$, $V_1$, and aVR, and the ratio of R and the S durations in I, II, $V_6$, and $V_7$. However, they found it difficult to supply threshold values for these parameters and were hesitant to indicate the number of leads in which these thresholds should be met for the diagnosis to be made. Because of the age-dependency of the QRS duration, as illustrated in Fig. 5.1, we defined the difference between the measured QRS duration and the upper limit of normal for age (ULN) as a new feature (ΔQRS duration), and specified a search range from -20 to 20 ms for EXPLORE. Furthermore, the wave-duration ratios of the RBBB cases were often seen to be larger than two. We then introduced another feature, #prolonged waves, indicating the number of leads in which this condition is met, and defined the following subrule:

$$(\Delta QRS\ duration > ?\ AND\ \#prolonged\ waves > ?) \tag{5.2}$$

Using the training set, EXLORE found that thresholds of 5 ms for ΔQRS duration and 3 for #prolonged waves yielded a rule which fulfilled our performance constraints. We decided to use the ULN for the QRS duration instead of the ULN + 5 ms, because the difference in performance proved to be negligible. This gave the following decision rule:

$$If\ (QRS\ duration > ULN\ AND\ \#prolonged\ waves > 3)\ then\ RBBB \tag{5.3}$$

This simple rule had a sensitivity of 86.4% and a specificity of 97.9% on the training set. On the test set, the sensitivity was 84.0% and the specificity 95.3%, which was considered acceptable.

In a similar way, we developed decision rules for all diagnostic categories and statements as shown in Table 5.2. In addition, PEDMEANS can give localization (e.g., pos-

terior, inferior), severity scores (e.g., marked, mild), and certainty scores (e.g., probable, possible) for some categories. Also, PEDMEANS can give an explanation of its reasoning by showing the major fulfilled criteria for each given statement.

**Table 5.2.** Diagnostic statements of PEDMEANS.

| Rhythm categories | Contour categories |
|---|---|
| aberrantly conducted complexes | axis deviation |
| accelerated AV junctional rhythm | biventricular hypertrophy |
| atrial fibrillation | high voltage |
| atrial tachycardia | infarct |
| AV junctional escapes | intraventricular conduction delay |
| AV junctional rhythm | left bundle branch block |
| AV junctional tachycardia | long QTc interval |
| ectopic atrial rhythm | low voltage |
| first degree AV block | left ventricular hypertrophy |
| idioventricular rhythm | left atrial hypertrophy |
| PR interval variation | right atrial hypertrophy |
| premature atrial complexes | right bundle branch block |
| premature supraventricular complexes | repolarization disturbance |
| premature ventricular complexes | right ventricular hypertrophy |
| second degree AV block | ST depression |
| short PR interval | ST elevation |
| sinus arrhythmia | unusual P axis |
| sinus bradycardia | Wolf-Parkinson-White syndrome |
| sinus rhythm | |
| sinus tachycardia | |
| supraventricular escapes | |
| supraventricular tachycardia | |
| ventricular escapes | |

In Table 5.3, the classification matrices for the test set are given for RBBB, LBBB, RVH and LVH. Table 5.4 gives the performance on the training and test sets.

**Table 5.3.** Classification matrices of right bundle branch block (RBBB), left bundle branch block (LBBB), right ventricular hypertrophy (RVH), left ventricular hypertrophy (LVH) for the test set.

| RBBB | Cardiologists | | | |
|---|---|---|---|---|
| PEDMEANS | absent | possible | probable | definite |
| absent | 557 | 7 | 6 | 2 |
| possible | 0 | 0 | 0 | 0 |
| probable | 18 | 2 | 5 | 8 |
| definite | 6 | 2 | 7 | 22 |

| LBBB | Cardiologists | | | |
|---|---|---|---|---|
| PEDMEANS | absent | possible | probable | definite |
| absent | 636 | 0 | 0 | 1 |
| possible | 0 | 0 | 0 | 0 |
| probable | 1 | 0 | 0 | 0 |
| definite | 1 | 0 | 0 | 3 |

| RVH | Cardiologists | | | |
|---|---|---|---|---|
| PEDMEANS | absent | possible | probable | definite |
| absent | 461 | 46 | 5 | 2 |
| possible | 23 | 14 | 6 | 6 |
| probable | 6 | 10 | 11 | 1 |
| definite | 2 | 6 | 18 | 25 |

| LVH | Cardiologists | | | |
|---|---|---|---|---|
| PEDMEANS | absent | possible | probable | definite |
| absent | 491 | 56 | 3 | 2 |
| possible | 23 | 11 | 2 | 0 |
| probable | 10 | 7 | 4 | 1 |
| definite | 4 | 6 | 12 | 10 |

**Table 5.4.** Performance measures of the training set and test set for the major diagnostic categories.

| | Training set | | Test set | |
|---|---|---|---|---|
| Abnormality | Sens (%) | Spec (%) | Sens (%) | Spec (%) |
| RVH | 71.7 | 97.1 | 74.3 | 96.5 |
| LVH | 80.0 | 95.0 | 79.4 | 95.6 |
| RBBB | 86.4 | 97.9 | 84.0 | 95.3 |
| LBBB | 62.5 | 99.3 | 75.0 | 99.7 |

## 5.4 Discussion

We developed and evaluated a computer program for interpretation of pediatric 12-lead ECGs with the use of a modified version of our EXPLORE induction algorithm. The combination of expertise of the pediatric cardiologist and the algorithm appeared to be a powerful tool in the development of the decision support system. By allowing more user-interaction we could incorporate the knowledge of the pediatric cardiologists, which decreased the search time of EXPLORE and improved the comprehensibility of the resulting decision rules. Furthermore, the ability of EXPLORE to optimize sensitivity with constraints on the specificity enabled us to develop rules that fulfilled our performance requirements. Since high specificities are generally considered mandatory for a rule to be practically useful, we chose to develop diagnostic criteria with specificities of at least 95%, while maintaining sensitivity as high as possible. With other induction techniques, such as CART [25], this would have been more difficult, since the user has no direct control over the sensitivity and specificity and must resort to a process of trial-and-error in deriving a classifier that suits his needs.

The collection of normal ECGs proved to be essential for obtaining reliable normal limits that could be used to define age-dependent criteria. During the expert interviews, it became clear that our pediatric cardiologists had difficulties to define age-dependent criteria, and even more difficulties to use these criteria consistently during their interpretations. Therefore, we estimated continuous age-dependent normal limits and used these as threshold values in the decision rules to avoid abrupt changes in diagnosis with small changes in age. These new normal limits differ substantially from those commonly used and suggest that diagnostic criteria for the pediatric ECG should be adjusted [5]. This presents a difficulty in program validation. Since the pediatric cardiologists were not able to use the new normal limits during their interpretations, differences with the interpretation program were unavoidable. For example for intraventricular conduction delay, in which the QRS duration is the main criterion, it was difficult to use the pediatric cardiologists as a reference. If PEDMEANS stated that the QRS was prolonged on basis of the new normal limits, the pediatric cardiologists had no ground to overrule this. Therefore, in this study we only present performance measures for a selected set of abnormalities. In might be interesting to have the pediatric cardiologists re-examine a set of ECGs with the newly estimated normal limits to see whether the disagreement between program and reference decreases.

Comparison of our results with those of earlier pediatric ECG computer programs is difficult: most studies did not report performance measures, sampling rates were much lower, and often the 12 leads were not recorded simultaneously. A validated

database of pediatric ECGs for testing and comparing pediatric ECG interpretation programs, following the lines of the seminal Common Standards for Quantitative Electrocardiography (CSE) study [26] for adult ECG programs, would be extremely helpful but is a huge task not easily accomplished. In our present evaluation, the performance on the test set was comparable with the performance on the training set. This indicates that the rules were not over-specialized on the training set. All specificities on the test set proved to be greater than 95% as we aimed for. Corresponding sensitivities, in this study population, were considered quite acceptable for use in a clinical setting.

## 5.5 Study limitations

We did not present performance measures for the rhythm part of PEDMEANS, because our dataset contained only few rhythm abnormalities to make an evaluation possible. Although, we believe that the modified adult rhythm classification is a good starting point, we should still collect a dataset that enables us to optimize and validate the rhythm section of PEDMEANS.

Our database of normal ECGS is limited in that it does not contain children younger than 11 days, which only allowed us to partially evaluate the performance. For this reason we choose not to make definite statements for children younger than 2 weeks, but only suggest possible abnormalities. A final limitation of our study is that the number of cases for some categories, (e.g., LBBB) is small and needs to be expanded to more reliably gauge the performance for these categories. Preferably, PEDMEANS should also be evaluated in different clinical centers.

## 5.6 Conclusions

A computer program for the interpretation of pediatric ECGs has been developed and evaluated. The performance of the program, on our study population, seems to justify its use in a clinical setting.

## 5.7 References

[1] Guller B, O'Brien PC, Smith RE, Weidman WH, DuShane JW. Computer interpretation of Frank vectorcardiogram in normal infant: Longitudinal and cross-sectional observations from birth to 2 years of age. J Electrocardiol 1975;8:201–8.

[2] Hamilton KH, Clifton JF, Laks MM. Computerized pediatric ECG interpretation. A review of the first year of live. In: Pryor A, Bailey J, editors. Computerized interpretation of the ECG IV. New York: Engineering Foundation, 1979; pp. 239–312.

[3] Brohet CR, Derwael-Barchy C, Robert A, Fesler R, Styns M, Brasseur LA, et al. Computer interpretation of pediatric Frank vectorcardiograms in the evaluation of congenital heart disease. Am J Cardiol 1983;52:127–32.

[4] Davignon A, Rautaharju P, Boisselle E, Soumis F, Megelas M, Choguette A. Normal ECG standards for infants and children. Pediatr Cardiol 1979/80;1:123–31.

[5] Rijnbeek PR, Witsenburg M, Schrama E, Hess J, Kors JA. New normal limits for the paediatric electrocardiogram. Eur Heart J 2001;22:702–11.

[6] Francis DB, Miller BL, Benson DW. A new computer program for the analysis of pediatric scalar electrocardiograms. Comput Biomed Res 1981;14:63–77.

[7] Laks MM. Eight years of experience in development of a pediatric computerized ECG program. In: Ripley K, editor. Computers in Cardiology. Los Alamitos, California: IEEE Computer Society, 1984; pp. 159–64.

[8] Laks MM. A computer program for interpretation of infant and children electrocardiogram. In: Willems J, van Bemmel J, Zywietz C, editors. Computer ECG Analysis: Towards Standardization. Amsterdam: North-Holland, 1986; pp. 59–65.

[9] Macfarlane PW, Coleman EN, Devine B, Houston A, McLaughlin S, Aitchison TC, et al. A new 12-lead pediatric ECG interpretation program. J Electrocardiol 1990;23:76–81.

[10] Zywietz C, Klusmeier S, Bernsau U. New VCG-analysis program for children with multivariate diagnostic classification. In: Ostrow H, Ripley K, editors. Computers in Cardiology. New York: IEEE, 1977; pp. 95–100.

[11] Guller B, Lau FY, Dunn RA, Pipberger HA, Pipberger HV. Computer analysis of changes in Frank vectorcardiograms of 666 normal infants in the first 72 hours of life. J Electrocardiol 1977;10:19–26.

[12] Michaelis J, Lippold R, Scheidt E, Jüngst B, Stopfkuchen H, Schönberger W. Automated analysis of vectorcardiograms in childhood for screening purposes. In: Macfarlane P, editor. Progress in Electrocardiology. Turnbridge Wells: Pitman Medical, 1979; pp. 271–6.

[13] Macfarlane PW, Coleman EN, Pomphrey EO, McLaughlin SC, Houston A, Aitchison T. Normal limits of the high-fidelity pediatric ECG. Preliminary observations. J Electrocardiol 1989;22:162–8.

[14] Golden DP, Wolthuis RA, Hoffler GW. A spectral analysis of the normal resting electrocardiogram. IEEE Trans Biomed Eng 1973;20:366–72.

[15] Barr RC, Spach MS. Sampling rates required for digital recording of intracellular and extracellular cardiac potentials. Circulation 1977;55:40–8.

[16] Garson A. Clinically significant differences between the "old" analog and the "new" digital electrocardiograms. Am Heart J 1987;114:194–7.

[17] Maroney M, Rantz LA. Electrocardiograms in 679 healthy infants and children. Pediatrics 1950;5:396.

[18] Ziegler RF. Electrocardiographic studies in normal infants and children. Springfield: Charles C. Thomas, 1951.

[19] Strong WB, Downs TD, Liebman J, Liebowitz R. The normal adolescent electrocardiogram. Am Heart J 1972;83:115–28.

[20] Liebman J, Plonsey R, Gillette PC. Pediatric cardiology. Baltimore: Williams & Wilkins, 1982.

[21] Macfarlane PW, McLaughlin SC, Devine B, Yang TF. Effects of age, sex, and race on ECG interval measurements. J Electrocardiol 1994;27:14–9.

[22] Van Bemmel JH, Kors JA, Van Herpen G. Methodology of the modular ECG analysis system MEANS. Methods Inf Med 1990;29:346–53.

[23] Kors JA, Hoffmann AL. Induction of decision rules that fulfil user-specified performance requirements. Pattern Recognition Letters 1997;19:1187–1195.

[24] Weiss S, Kulikowski C. Computer Systems that Learn: Classification and Prediction Methods from Statistics, Neural Nets, Machine Learning and Expert Systems. San Mateo, CA: Morgan Kaufmann, 1990.

[25] Breiman L, Friedman J. Classification and regression trees. Belmont, CA: Wadsworth and Brooks, 1984.

[26] Willems JL, Arnaud P, van Bemmel JH, Degani R, Macfarlane PW, Zywietz C. Common standards for quantitative electrocardiography: goals and main results. CSE Working Party. Methods Inf Med 1990;29:263–71.

# 6 Electrocardiographic criteria for left ventricular hypertrophy in children

P.R. Rijnbeek*, G. van Herpen*, M. Witsenburg†, A.D.J. ten Harkel†, L. Kapusta‡, J.A. Kors*

* Dept. of Medical Informatics, Erasmus University Medical Center, Rotterdam, The Netherlands
† Dept. of Pediatric Cardiology, Erasmus University Medical Center, Rotterdam, The Netherlands
‡ Children's Heart Center, Radboud University Medical Center, Nijmegen,The Netherlands

Submitted for publication

# Abstract

In pediatrics, the electrocardiogram (ECG) is still an important tool for the initial diagnosis of left ventricular hypertrophy (LVH). Earlier studies in patient groups with specific cardiac diseases report sensitivities varying widely from 20% to 67% for detecting LVH. A validity study on an unselected pediatric hospital population with mixed cardiac abnormalities has not been performed yet. Futhermore, several criteria that were shown to improve the detection of LVH in adults have not been tested in children.

The study population consisted of 832 children from whom a 12-lead electrocardiogram and an M-mode echocardiogram were taken on the same day at the Sophia Children's Hospital in Rotterdam in the period 2003 to 2005. Two pediatric cardiologists were independently presented with the medical record of each patient and scored the likelihood of volume or pressure overload of the left and right ventricle on a three-point scale (absent, possible, probable). Disagreements, defined as more than one point difference were solved by consensus. LVH in the echocardiogram was defined as a LVM index ($LVM_I$), taken as LVM (in g) divided by body weight (in kg), exceeding a gender-independent value of 3. The validity of current pediatric and adult ECG criteria was judged on the basis of an abnormal $LVM_I$, alone and in combination with the clinical LVH score of the cardiologists. In addition, a search was made for possibly better performing combinations of criteria with the use of automatic learning techniques.

The ECG criteria have relatively low (<25%) sensitivity rates at 95% specificity in the pediatric hospital population when an elevated $LVM_I$ is taken as the reference for LVH. The performance appears to improve with age. When clinical evidence is also taken into account in the definition of LVH, the sensitivity improved to 43%. By combining two ECG parameters the maximum sensitivity is raised to 32% in the group with abnormal $LVM_I$ and to 51% in the group with a combination of abnormal $LVM_I$ and clinical evidence suggestive of LVH.

The sensitivity of the pediatric ECG in detecting LVH in a hospital population is low when $LVM_I$ is taken as the reference. When the clinical evidence of LVH is also taken into account, the performance of the ECG is considerably higher. The sensitivity can be further improved if ECG parameters are combined.


**Keywords:** electrocardiography, pediatrics, LVH, echocardiography

## 6.1 Introduction

Left ventricular hypertrophy (LVH) results from adaptation of the heart to increased hemodynamic burden. Therefore, early detection of LVH is important, especially in children. Although the 12-lead electrocardiogram (ECG) is still valued as an initial diagnostic test for LVH, its sensitivity in this respect leaves to be desired. In a recent large study on HIV-infected children, Rivenes et al. [1] found sensitivities of <20% at specificity levels of 88% to 92%. On the other hand, in a number of smaller studies on patients with a specific cardiac disease the performance of the pediatric ECG was found to be higher [2–6]. For example, Fogel et al. [2] found in a group of 19 aortic stenosis patients and 21 normals a sensitivity of 67% at 95% specificity. None of these studies was done on an unselected pediatric hospital population with mixed cardiac abnormalities. For the present investigation we have collected such a population and have sought to improve the sensitivity of the pediatric ECG in detecting LVH by means that issue from the following considerations:

Firstly, in pediatric electrocardiology only a limited number of criteria have been used for assessing LVH. Several criteria that were shown to improve the detection of LVH in adults have not been tested in children. Pediatric electrocardiographers have focused on the QRS amplitude, although the time-voltage area of the QRS complex or its approximation by the product of maximum QRS voltage and QRS duration were proposed as useful criteria to improve LVH diagnosis in adults [7]. Likewise, in adults combinations of criteria have been shown to improve performance [8–10], which approach has not yet been attempted for children.

Secondly, older validity studies have used obsolete limits of normal for the pediatric ECG parameters. In a previous study [11], we established new age-dependent normal limits that differ considerably from the older figures [12,13]. Hitherto, normal limits have not been revised for LVH detection.

Finally, the reference standard for LVH on the ECG has usually been the left ventricular mass (LVM) as estimated from echocardiographic measurements. This reference has been criticized for being vulnerable to measurement error and for oversimplifying the geometry of the left ventricle [14,15]. Alternatively, a combination of increased LVM and clinical evidence of volume or pressure overload of the left ventricle may be a better reference standard for the validity of ECG criteria.

## 6.2 Methods

### Study population

We collected data of all 904 children from whom a 12-lead electrocardiogram and an echocardiogram were taken on the same day at the Sophia Children's Hospital in Rotterdam in the period 2003 to 2005. We excluded 5 children who had received heart transplantation and 67 children with complete transposition of the great arteries, leaving a study population of 832 children. In Table 6.1, the age and sex distribution of the population is shown.

**Table 6.1.** Age and sex distribution of the study population.

| Age | Male | Female | Total |
|---|---|---|---|
| 0 - 2 months | 14 | 18 | 32 |
| 3 - 5 months | 15 | 8 | 23 |
| 6 - 11 months | 18 | 13 | 31 |
| 1 - 2 years | 50 | 45 | 95 |
| 3 - 4 years | 66 | 47 | 113 |
| 5 - 7 years | 77 | 65 | 142 |
| 8 - 11 years | 106 | 82 | 188 |
| 12 - 15 years | 109 | 99 | 208 |
| Total | 455 | 377 | 832 |

Clinical data were studied by two pediatric cardiologists (MW and ADJH) who were independently presented with the medical record of each patient. Each cardiologist had to score the likelihood of volume or pressure overload of the right and/or left ventricle on a three-point scale (0=absent, 1=possible, 2=probable). LVH was considered present in patients with, e.g., aortic stenosis or regurgitation, and RVH in patients with tetralogy of Fallot or pulmonary hypertension. Disagreements, defined as more than one point scoring difference, were settled by consensus.

### Electrocardiography

Twelve-lead ECGs were recorded by means of a PC-based acquisition system (Welch Allyn Cardio Control, Delft, The Netherlands) at a sampling rate of 1200 Hz. Following common practice in the department of pediatric cardiology in Rotterdam, $V_3$ was

moved to the $V_{3R}$ position, and $V_5$ was moved to the $V_7$ position. All ECGs were processed by the pediatric ECG computer program PEDMEANS [16]. To reduce noise, PEDMEANS computes a representative averaged beat for each of the twelve leads, from which ECG measurements are derived. Wave onsets and offsets as found by PEDMEANS were visually checked.

A total of 15 ECG parameters for diagnosing LVH in children were evaluated (Table 6.2). In addition to the standard pediatric parameters, we included parameters based on $V_{3R}$ and $V_7$, and defined an additional version of the Sokolow-Lyon criterion by using $V_{3R}$ and $V_7$ instead of $V_1$ and $V_6$. Furthermore, the sum of the R- and S-wave amplitudes in all leads (12-lead sum), and the QRS voltage-duration product and voltage-time integral versions of the Sokolow-Lyon and 12-lead sum criteria were added. The sensitivities of the voltage-duration product version and the voltage-time integral version were compared with their amplitude versions by means of the McNemar's modification of the chi-square method for paired proportions.

**Table 6.2.** Evaluated ECG parameters for LVH.

| Parameter | Description |
|---|---|
| $SV_{3R}$ | S-wave amplitude in $V_{3R}$ |
| $SV_1$ | S-wave amplitude in $V_1$ |
| $RV_6$ | R-wave amplitude in $V_6$ |
| $RV_7$ | R-wave amplitude in $V_7$ |
| $TV_6$ | inverted T wave in $V_6$ |
| $TV_7$ | inverted T wave in $V_7$ |
| $SV_1 + RV_6$ | Sokolow-Lyon voltage |
| $(SV_1 + RV_6) \times QRS_d$ | Sokolow-Lyon voltage-duration product |
| $(SV_1 + RV_6)$ area | Sokolow-Lyon voltage-time integral |
| $SV_{3R} + RV_7$ | additional Sokolow-Lyon voltage |
| $(SV_{3R} + RV_7) \times QRS_d$ | additional Sokolow-Lyon voltage-duration product |
| $(SV_{3R} + RV_7)$ area | additional Sokolow-Lyon voltage-time integral |
| 12-lead sum | sum of top-top deflections of all leads |
| 12-lead sum $\times$ $QRS_d$ | 12-lead sum voltage-duration product |
| 12-lead sum area | 12-lead sum voltage-time integral |

The polarity of the T wave in $V_6$ or $V_7$ was taken as a binary value. For all other parameters, reference values were derived from a normal population of 1912 children aged 11 days to 16 years [11]. The 98th percentile of the parameter distribution was taken as the upper limit of normal (ULN). We estimated age-dependent percentile

curves using a two-stage parametric approach described before [11]. As an illustration, the age-dependent normal curve for the 12-lead sum is presented in Fig. 6.1. For each LVH parameter, the difference between the parameter value and its corresponding normal limit was taken for further processing.

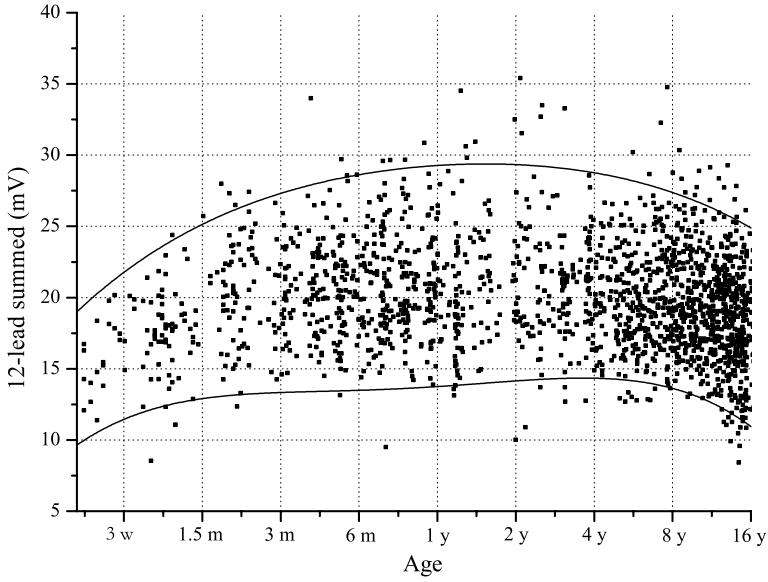


**Figure 6.1.** Scatter diagram for the 12-lead sum against age. The solid lines indicate the $2^{nd}$ and $98^{th}$ percentile curves.

### Echocardiography

The echocardiograms were recorded with a Philips Sonos 5500 (Philips, Best, The Netherlands). The following M-mode measurements were obtained according to the American Society of Echocardiographers' convention [17]: interventricular septum thickness at end diastole (IVSd), posterior wall thickness at end diastole (PWLVd), and left ventricular internal dimension at end diastole (LVIDd). Left ventricular mass (LVM) was calculated with the anatomically validated formula of Devereux [18]:

$$LVM = 0.8(1.04(IVSd + PWLVd + LVIDd)^3 - (LVIDd)^3) + 0.6\ [g]$$

The LVM index ($LVM_I$) adjusts for body size and is taken as LVM (in g) divided by body weight (in kg). As shown in Fig. 6.2, we estimated a partition value of 3 for

$LVM_I$, based on a set of 587 normal echocardiograms of 361 boys and 226 girls aged birth to 18 years previously described by Overbeek et al. [19]. We calculated the LVM using the formula of Devereux from the measurements provided to us by Overbeek.

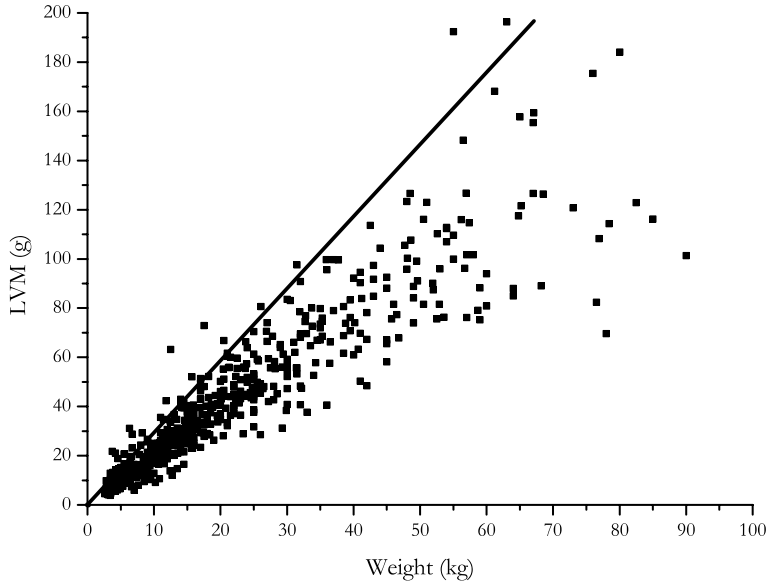


**Figure 6.2.** Left ventricular mass (LVM) calculated on 587 normal echocardiograms plotted against body weight. The solid line represents the partition value of the LVM index, $LVM_I$ = LVM/Weight = 3.

### Reference standards for LVH

The ECG criteria for LVH were validated using two different reference standards. Like in other validation studies, one standard is based only on the $LVM_I$. LVH was considered present if $LVM_I > 3$ (LVM+). In the second standard the clinical evidence of LVH is also taken into consideration and LVH was defined at two different likelihood levels. LVH was assumed to be present at the first level, if $LVM_I > 3$ and at least one cardiologist scored "possible" (LVM+L1), and at the second level if $LVM_I > 3$ and at least one cardiologist scored "probable" (LVM+L2). Note that the difference between the two cardiologists can never be higher than one scoring point due to the obtained consensus. The cases with a $LVM_I \leq 3$ were classified as non-LVH.

### Combination of ECG parameters

Combinations of LVH-ECG parameters were evaluated using our EXPLORE induction algorithm [20]. EXPLORE can search for the decision rule that has the highest sensitivity at a user-specified level of specificity. In this study, we wanted EXPLORE to find the best decision rule consisting of two ECG parameters at a specificity of 95%. EXPLORE takes as its input a set of ECGs, each presented by a set of parameters, and a label indicating whether LVH is present of not. We searched for the best parameter combination for each of the three LVH definitions: LVM+, LVM+L1, and LVM+L2.

## 6.3 Results

### Performance of individual LVH-ECG parameters

Table 6.3 shows the diagnostic performance of all ECG parameters for different age groups and for the total study population, taking LVM+ as LVH definition.

**Table 6.3.** Sensitivity (in %) at 95% specificity for three age groups and for the total population taking LVM+ as LVH definition.

| Parameter | 0 - 5 yr | 6 - 11 yr | 12 - 15 yr | 0 - 15 yr |
|---|---|---|---|---|
| $SV_1$ | 16.5 | 17.1 | 19.4 | 16.7 |
| $SV_{3R}$ | 18.3 | 20.0 | 33.3 | 20.4 |
| $RV_6$ | 12.2 | 15.7 | 16.7 | 16.3 |
| $RV_7$ | 19.1 | 20.0 | 25.0 | 22.2 |
| $TV_6$ | 7.0 | 10.0 | 13.9 | 9.0 |
| $TV_7$ | 17.4 | 14.3 | 25.0 | 15.8 |
| $SV_1 + RV_6$ | 20.0 | 22.9 | 19.4 | 22.2 |
| $(SV_1 + RV_6) \times QRS_d$ | 16.5 | 21.4 | 22.2 | 24.0 |
| $(SV_1 + RV_6)$ area | 19.1 | 21.4 | 19.4 | 24.4 |
| $(SV_{3R} + RV_7)$ | 29.6 | 18.6 | 27.8 | 25.3 |
| $(SV_{3R} + RV_7) \times QRS_d$ | 22.6 | 20.0 | 22.2 | 23.1 |
| $(SV_{3R} + RV_7)$ area | 20.0 | 20.0 | 19.4 | 20.8 |
| 12-lead sum | 15.7 | 17.1 | 22.2 | 17.2 |
| 12-lead sum $\times QRS_d$ | 19.1 | 17.1 | 22.2 | 18.6 |
| 12-lead sum area | 22.6 | 14.3 | 27.8 | 22.6 |
| $\#cases\ (LVH+/LVH-)$ | 115/225 | 70/214 | 36/172 | 221/611 |

In the complete set (0-15 years), the additional version of the Sokolow-Lyon parameter $(SV_{3R} + RV_7)$ performs best, with 25.3% sensitivity. Multiplying by QRS duration or

taking the voltage-time integral does not improve this criterion. However, there is an increase in sensitivity of 31% of the area version compared to the amplitude-only version of the 12-lead sum in the complete set (p=0.045).

The effect of age on the results is mixed. For different age groups different parameters perform best and the improvement of the 12-lead sum is not consistent in all age groups. However, this may be accounted for by the lower number of LVH cases in the higher age groups. For $SV_1$, $SV_{3R}$, $RV_6$, $RV_7$, $TV_6$, and $(SV_1 + RV_6)\times$ QRS duration an increasing performance is seen with age. In all age groups, $SV_{3R}$ appears to perform better than $SV_1$, and $RV_7$ better than $RV_6$.

In Table 6.4 the diagnostic performance of the individual ECG parameters is presented for the different definitions of LVH.

**Table 6.4.** Sensitivity (in %) at 95% specificity for three LVH definitions.

| Parameter | LVM+ | LVM+L1 | LVM+L2 |
|---|---|---|---|
| $SV_1$ | 16.7 | 21.1 | 30.0 |
| $SV_{3R}$ | 20.4 | 25.9 | 33.8 |
| $RV_6$ | 16.3 | 19.0 | 26.3 |
| $RV_7$ | 22.2 | 25.9 | 36.3 |
| $TV_6$ | 9.0 | 10.9 | 12.5 |
| $TV_7$ | 15.8 | 20.4 | 26.3 |
| $SV_1 + RV_6$ | 22.2 | 25.2 | 35.0 |
| $(SV_1 + RV_6) \times QRS_d$ | 24.0 | 29.9 | 31.3 |
| $(SV_1 + RV_6)$ area | 24.4 | 30.6 | 41.3 |
| $(SV_{3R} + RV_7)$ | 25.3 | 30.6 | 42.5 |
| $(SV_{3R} + RV_7) \times QRS_d$ | 23.1 | 28.6 | 37.5 |
| $(SV_{3R} + RV_7)$ area | 20.8 | 26.5 | 37.5 |
| 12-lead sum | 17.2 | 21.1 | 33.8 |
| 12-lead sum $\times QRS_d$ | 18.6 | 21.8 | 33.8 |
| 12-lead sum area | 22.6 | 27.9 | 37.5 |
| $\#cases\ (LVH+/LVH-)$ | 221/611 | 147/611 | 80/611 |

The sensitivity of all ECG criteria improves considerably when, apart from LVM+, the clinical evidence of LVH is also required in the definition of LVH. The more certain the cardiologists were about the presence of LVH, the better the ECG performs. The additional Sokolow-Lyon criterion $(SV_{3R} + RV_7)$ performs best for all reference standards. Removal of all LVH cases from the LVM+L2 group which were also marked by at least one cardiologist as probable RVH (n=15) did hardly change the performance of the parameters.

**Performance of a combination of two LVH-ECG parameters**

Using the EXPLORE algorithm, we found that for all definitions of LVH the best rule consisted of $(SV_{3R} + RV_7)$ area in conjunction with 12-lead sum area. The threshold values of these parameters were optimized by EXPLORE to ensure 95% specificity. Table 6.5 shows that this combination improves the sensitivity by 21-25% as compared to the best single parameter $(SV_{3R} + RV_7)$.

**Table 6.5.** Sensitivities (in %) for the best single parameter $(SV_{3R} + RV_7)$ and the best combination of two parameters ($(SV_{3R} + RV_7)$ area with 12-lead sum area) for the three LVH definitions at a 95% specificity level.

| LVH definition | 1 parameter | 2 parameters | % increase |
|---|---|---|---|
| LVM+ | 25.3 | 31.7 | 25.3 |
| LVM+L1 | 30.6 | 38.1 | 24.5 |
| LVM+L2 | 42.5 | 51.3 | 20.7 |

## 6.4 Discussion

We performed the first large study on the validity of ECG criteria for detecting LVH in an unselected pediatric hospital population with up-to-date reference standards for both electrocardiographic and echocardiographic parameters. The unreliability of reference standards for both the ECG criteria and the echocardiographic based criteria may previously have accounted for the disappointing performance of the ECG. Rivenes et al. [1] demonstrated that the often used standard of normal ECG limits by Davignon et al. [12] needed to be revised. They derived new age-adjusted reference values for their HIV-uninfected group of children below 6 years. With these new normal values the sensitivity slightly decreased (from <20% to <17%) while specificity improved (from 88%-92% to 94%-100%). Regarding the echocardiogram, a recent study provided echocardiographic dimensions that differed significantly from previous data, probably owing to improved measurement techniques [19].

The sensitivity of the ECG for detecting LVH in children is relatively low when LVM+ is taken as LVH definition (Table 6.3). The best single parameter is $(SV_{3R} + RV_7)$ with a sensitivity of 25.3% at 95% specificity. Notably, $SV_{3R}$, $RV_7$ and $(SV_{3R} + RV_7)$ perform better than $SV_1$, $RV_6$, and $(SV_1 + RV_6)$, respectively. The best performing parameter that uses leads from the standard 12-lead ECG is the voltage-time integral

version of the Sokolow-Lyon criterion with a sensitivity of 24.4%. Further, the use of voltage-duration product and voltage-time integral of the 12-lead sum resulted in higher sensitivities than the amplitude-only version in the total population. However, overall the use of voltage-duration products and voltage-time integrals for the diagnosis of LVH remains less effective than in adults.

In most studies on the usefulness of the ECG in detecting LVH, echocardiographic determined LVM was taken as the reference standard. However, we believe that the ECG should not only be judged by taking LVM as the reference. LVM determination suffers from large measurement errors [14,15]. When LVM is overestimated the ECG is penalized for stating normal. We therefore also used combinations of increased LVM and clinical indices of volume or pressure overload to ascertain the validity of ECG criteria. We found that the ECG is performing considerably better when there is more certainty about the presence of LVH on clinical grounds. Fogel et al. [2] showed that the sensitivity was 67% in a small group of 19 children with aortic stenosis. Unfortunately, we could not collect a large enough group of aortic stenosis patients without other abnormalities to make a better comparison with Fogel possible.

In adults combining different ECG-LVH parameters is known to improve performance [8–10]. The combination of highly specific criteria can be a fruitful approach to increase sensitivity without generating an unacceptable number of false-positives. We restricted the search of our EXPLORE algorithm to combinations of only two parameters because more parameters might introduce over-specialization on a data set of our size. A combination of $(SV_{3R} + RV_7)$ area and 12-lead sum area was found to improve the sensitivity by 25% in the LVM+ and LVM+L1 groups and by 21% in the LVM+L2 group, at 95% specificity. Combination of parameters can thus be useful in the pediatric population also.

Our study has several limitations. Firstly, we validated the ECG criteria for LVH using LVM as measured by M-mode echocardiography. However, there are studies in adults that suggest that cardiac MRI can more accurately and reproducibly measure LVM [14,21]. In children this has not yet been confirmed. Secondly, in our study $V_3$ was moved to the $V_{3R}$ position, and $V_5$ was moved to the $V_7$ position. Therefore, the 12-lead sum included these leads instead of $V_3$ and $V_5$, which makes comparison with other studies using standard lead sets more difficult. Thirdly, we could not determine the effect of concomitant RVH on the performance of the ECG in detecting LVH due to the low number of cases scored as probable RVH by the cardiologists (n=15). Fourthly, we could not establish the effect of gender because of the limited size of the study population. In adults gender influences the performance of the ECG criteria for LVH [14,22,23]. In our normal pediatric ECG population we did find gender differences in

QRS duration for all ages and in R- and S-wave amplitudes for children aged 12 to 16 years [11]. The effect of gender on ECG-LVH criteria in the pediatric population awaits investigation on a larger population.

## 6.5 Conclusion

The sensitivity of the pediatric ECG in detecting LVH is low when $LVM_I$ is taken as the reference. When the clinical evidence of LVH is also taken into account, the performance of the ECG is considerably higher. The sensitivity can be further improved if ECG parameters are combined.

## 6.6 References

[1] Rivenes SM, Colan SD, Easley KA, Kaplan S, Jenkins KJ, Khan MN, et al. Usefulness of the pediatric electrocardiogram in detecting left ventricular hypertrophy: results from the Prospective Pediatric Pulmonary and Cardiovascular Complications of Vertically Transmitted HIV Infection (P2C2 HIV) multicenter study. Am Heart J 2003;145:716–23.

[2] Fogel MA, Lieb DR, Seliem MA. Validity of electrocardiographic criteria for left ventricular hypertrophy in children with pressure- or volume-loaded ventricles: comparison with echocardiographic left ventricular muscle mass. Pediatr Cardiol 1995;16:261–9.

[3] Sastroasmoro S, Madiyono B, Oesman IN. Sensitivity and specificity of electrocardiographic criteria for left ventricular hypertrophy in children with rheumatic heart disease. Paediatr Indones 1991;31:233–44.

[4] Panza JA, Maron BJ. Relation of electrocardiographic abnormalities to evolving left ventricular hypertrophy in hypertrophic cardiomyopathy during childhood. Am J Cardiol 1989;63:1258–65.

[5] Louie EK, Maron BJ. Hypertrophic cardiomyopathy with extreme increase in left ventricular wall thickness: functional and morphologic features and clinical significance. J Am Coll Cardiol 1986;8:57–65.

[6] Oda T, Hamamoto K, Morinaga H. Left ventricular hypertrophy in non-rheumatic myocarditis in children. Jpn Circ J 1982;46:1235–8.

[7] Okin PM, Roman MJ, Devereux RB, Pickering TG, Borer JS, Kligfield P. Time-voltage QRS area of the 12-lead electrocardiogram: detection of left ventricular hypertrophy. Hypertension 1998;31:937–42.

[8] Schillaci G, Verdecchia P, Borgioni C, Ciucci A, Guerrieri M, Zampi I, et al. Improved electrocardiographic diagnosis of left ventricular hypertrophy. Am J Cardiol 1994;74:714–9.

[9] Salles G, Leocadio S, Bloch K, Nogueira AR, Muxfeldt E. Combined QT interval and voltage criteria improve left ventricular hypertrophy detection in resistant hypertension. Hypertension 2005;46:1207–12.

[10] Warner RA, Ariel Y, Gasperina MD, Okin PM. Improved electrocardiographic detection of left ventricular hypertrophy. J Electrocardiol 2002;35 Suppl:111–5.

[11] Rijnbeek PR, Witsenburg M, Schrama E, Hess J, Kors JA. New normal limits for the paediatric electrocardiogram. Eur Heart J 2001;22:702–11.

[12] Davignon A, Rautaharju P, Boisselle E, Soumis F, Megelas M, Choguette A. Normal ECG standards for infants and children. Pediatr Cardiol 1979/80;pp. 123–31.

[13] Dickinson DF. The normal ECG in childhood and adolescence. Heart 2005;91:1626–30.

[14] Alfakih K, Walters K, Jones T, Ridgway J, Hall AS, Sivananthan M. New gender-specific partition values for ECG criteria of left ventricular hypertrophy: recalibration against cardiac MRI. Hypertension 2004;44:175–9.

[15] Crow RS, Hannan P, Grandits G, Liebson P. Is the echocardiogram an appropriate ECG validity standard for the detection and change in left ventricular size? J Electrocardiol 1996;29 Suppl:248–55.

[16] Rijnbeek PR, Witsenburg M, Szatmari A, Hess J, Kors JA. PEDMEANS: a computer program for the interpretation of pediatric electrocardiograms. J Electrocardiol 2001;34 Suppl:85–91.

[17] Lang RM, Bierig M, Devereux RB, Flachskampf FA, Foster E, Pellikka PA, et al. Recommendations for chamber quantification: a report from the American Society of Echocardiography's Guidelines and Standards Committee and the Chamber Quantification Writing Group, developed in conjunction with the European Association of Echocardiography, a branch of the European Society of Cardiology. J Am Soc Echocardiogr 2005;18:1440–63.

[18] Devereux RB, Alonso DR, Lutas EM, Gottlieb GJ, Campo E, Sachs I, et al. Echocardiographic assessment of left ventricular hypertrophy: comparison to necropsy findings. Am J Cardiol 1986;57:450–8.

[19] Overbeek LI, Kapusta L, Peer PG, de Korte CL, Thijssen JM, Daniels O. New reference values for echocardiographic dimensions of healthy Dutch children. Eur J Echocardiogr 2005;.

[20] Kors JA, Hoffmann AL. Induction of decision rules that fulfil user-specified performance requirements. Pattern Recognition Letters 1997;19:1187–95.

[21] Myerson SG, Montgomery HE, World MJ, Pennell DJ. Left ventricular mass: reliability of M-mode and 2-dimensional echocardiographic formulas. Hypertension 2002;40:673–8.

[22] Okin PM, Roman MJ, Devereux RB, Kligfield P. Gender differences and the electrocardiogram in left ventricular hypertrophy. Hypertension 1995;25:242–9.

[23] Casale PN, Devereux RB, Alonso DR, Campo E, Kligfield P. Improved sex-specific criteria of left ventricular hypertrophy for clinical and computer interpretation of electrocardiograms: validation with autopsy findings. Circulation 1987;75:565–72.

# 7 General Discussion

For both the general pediatrician and the pediatric cardiologist the electrocardiogram remains an important tool for the evaluation of children with possible cardiac disease. It is the cornerstone for interpretation of rhythm and conduction disorders, including QT prolongation syndromes. For recognition of either congenital or acquired structural heart disease it is routinely used in combination with physical examination. Based on clinical findings and the ECG those patients are selected in whom additional echocardiography is needed to confirm or exclude a cardiac malformation.

The wide variation in normal voltage range and the changes during growth hinder rapid visual interpretation of a pediatric ECG. This is especially true for the general pediatric practice where cardiac problems form a minority of daily care. Most pediatric cardiologists have been trained in ECG interpretation based on ECG criteria from several decades ago and will find it difficult to include changed criteria in visual ECG interpretation. Therefore establishment of new normal values and automatic ECG interpretation based on these new pediatric data are of great value for daily clinical practice.

In this thesis we described a number of issues related to the development of the PEDiatric Modular ECG Analysis System (PEDMEANS) for the automatic interpretation of pediatric ECGs. We will discuss our main findings and for each issue indicate further directions of research.

## Minimum bandwidth requirements

In Chapter 2, we established minimum bandwidth requirements for recording pediatric ECGs and recommended a minimum bandwidth of 250 Hz. This bandwidth is considerably higher than the previous recommendation of 150 Hz of the American Heart Association [1]. This finds its reason in the fact that in previous studies that determined the frequency content of the ECG, the study population was too small or the sampling frequency used by the recording system too low. Furthermore, we showed that the bandwidth is age-dependent, e.g., for children of 12 to 16 years the system bandwidth can be reduced to 150 Hz without introducing any performance degradation. To our knowledge, this is the first study that demonstrates the effect of bandwidth limitations for all 12 leads of the ECG and illustrates the influence of age on the frequency content. In the upcoming version of the recommendations of the AHA the minimum bandwidth requirements for recording pediatric ECGs will be modified on the basis of our study.

To determine the minimum bandwidth of the pediatric ECG we measured the absolute errors in maximum QRS amplitude for each simulated bandwidth, and de-

termined the percentage of records with an error higher than 25 $\mu$V. However, the influence of reduced bandwidth of ECG recording equipment on diagnostic interpretation has not been studied yet. The 25 $\mu$V threshold was only chosen because it was considered an amplitude difference still distinguishable by human interpreters from standard paper ECG recordings being equivalent to 1/4 mm.

## Normal limits

In Chapter 3, we derived new normal limits for the pediatric ECG that showed clinically significant differences with those in older studies. This implies that diagnostic criteria for the pediatric ECG should be adjusted. Recently, this was underlined by Dickinson who compared our data with the data of Davignon [2]. The number of studies that use the new normal limits is increasing [3–9]. Our normal curves demonstrated a strong age-dependency. Memorizing these normal limits is cumbersome for the pediatric cardiologist and automatic interpretation can thus be of great help. Possibly, the normal curves can be made available for use in a PDA to allow a quick reference during manual interpretations also.

We estimated continuous age-dependent normal limits to obtain a smooth transition of the normal limits with age. This reduces the effect of sudden jumps in diagnostic outcome of an automatic interpretation. With the use of tabulated normal limits a small change in age could result in a sudden change in normal limit.

Some aspects still need further consideration. The effect of electrode placement should be addressed. Earlier studies showed that large amplitude changes can occur if electrodes are shifted [10,11]. This must be an important source of variability especially in small children. Furthermore, in our study we used lead $V_{3R}$ and $V_7$ instead of leads $V_3$ and $V_5$ because these are easier to position on the chest in infants and young children. Normal limits for leads $V_3$ and $V_5$ need to be estimated in a future study. Moreover, additional data is needed for children aged 0 to 1 month which are low in number in our population.

## Formalization of diagnostic criteria

The performance of a decision support systems strongly depends on the expert knowledge that is implemented in the system. To assist in formalizing this knowledge, exhaustive rule generation proved to be of value. However, exhaustive generation of decision rules is very computer intensive. In Chapter 4, we presented a new version of our Exhaustive Procedure for LOgic-Rule Extraction (EXPLORE) algorithm and

showed that, with new insights, we could reduce the execution time of the algorithm considerably. Firstly, by instantiating feature-operator pairs instead of features and subsequently operators, the rules can be generated in a more efficient and systematic manner. Secondly, applying the subsumption principle reduced the number of candidate thresholds by 34% on average, compared to when only boundaries are taken, in the five data sets we used. Thirdly, the use of the branch and bound technique to generate rules exhaustively, by applying constantly updated bounds on the number of correct positives and false positives, has a very large favorable effect on the execution time. In this way the speed of induction is increased considerably, but heuristic approaches to induce larger rules or induce rules on large data sets will still be needed because of the exponential growth of the search space. Fourthly, an important asset of EXPLORE is the possibility to incorporate prior expert knowledge. This improves the comprehensibility of resulting rule, and may greatly speed up the induction process.

For the development of PEDMEANS we used a learning set of annotated ECGs and we tried to maximize the sensitivity of the decision rules while keeping specificity at $\geq$ 95%, a level that we considered necessary for the rules to be of practical utility. This was accomplished by an iterative procedure in which a pediatric cardiologist was asked to articulate his knowledge in the EXPLORE formalism. EXPLORE generated the best decision rule, and the comprehensibility and performance of this rule was discussed with the expert, possibly resulting in refinement of his specified prior knowledge. These steps could be reiterated until a satisfactory rule was obtained. We experienced that the pediatric cardiologists were often able to define rules which seemed to perform well on the easy cases. They could define the important parameters but the threshold values to be used were often more difficult to define accurately, probably due to the strong age-dependency of the normal limits. The combination of expert interviews and automatic learning techniques proved to be of great value in developing a decision support system like PEDMEANS. The cardiologist was allowed to define which parameters were thought to be of definite value and could suggest subrules to be optimized further by EXPLORE. In this way we obtained a better balance between performance and comprehensibility of the resulting classifier.

## Performance of PEDMEANS

Although the performance of PEDMEANS is acceptable for use in a clinical environment, the human expert is in some cases still a better ECG reader. This can partly be contributed to the fact that humans still outperform computer algorithms in recognizing waveforms, in particular P waves. Furthermore, the experts use insight in the underlying physical processes that result in an ECG on the body surface, which

is not always incorporated in the decision rules in the system. As discussed in Chapter 5, we were able to fine-tune the rules defined by the cardiologist with the use of EXPLORE and obtained high sensitivity values for most abnormalities. However, some remarks need to be made. Firstly, our reference standard consisted of a consensus of three cardiologists who were also involved in system development. An independent evaluation of the system by other pediatric cardiologists is therefore also advisable. Secondly, the cardiologists experienced difficulties in using correct age-dependent threshold values in their ECG criteria. During the expert interviews we found cases in which PEDMEANS scored abnormal based on the new normal limits even though the experts scored this case as normal. In such cases it is not clear what should be the reference, the computer or the human interpreter. Probably, if we would have given the cardiologists a measurement table for each ECG showing which parameters were abnormal on the basis of the new normal limits, the agreement between PEDMEANS and the cardiologist would improve. Thirdly, we might evaluate the system on ECG-independent data like the echo-LVH dataset collected in Chapter 6. However, our first goal was to develop a system that performs as well as an experienced group of pediatric cardiologists reading the ECG manually. Fourthly, the rhythm statements made by PEDMEANS can be further optimized and should be evaluated on a large database of cases with rhythm abnormalities. In the current version of PEDMEANS we only modified the criteria of the adult MEANS version using the new normal limits for children. Finally, it should be realized that the performance of decision support systems is dependent on the composition of the database used for development and validation purposes. Therefore, conclusions from our performance evaluation may not be applicable to other clinical environments.

## Development of new diagnostic criteria

In Chapter 6, we studied the validity of ECG criteria in detecting LVH based on echocardiographically determined left ventricular mass. We showed that current LVH criteria have relatively low (<25%) sensitivity rates at 95% specificity in the pediatric hospital population. When there is more certainty about the presence of LVH on clinical grounds the ECG is performing considerably better. Furthermore, by combining the $(SV_{3R} + RV_7)$ area with 12-lead sum area, with the use of the EXPLORE algorithm, the maximum sensitivities could be raised considerably. These combined criteria are possibly too complex for manual ECG interpretation but are easily implemented in a computer program. It should be interesting to follow this approach for other abnormalities as well, taking ECG independent material as the reference, e.g. for RVH.

### Future directions

To further improve the performance of PEDMEANS, a large data set of ECGs from children with a diversity of cardiac malformations, including rhythm disturbances, is needed. Ideally, a large database should become available for validation of pediatric ECG analysis systems, just like in the CSE study for adults [12].

Furthermore, in adults we were able to improve the performance of MEANS by combining ECG and VCG criteria [13]. Interestingly, to avoid the necessity of recording a VCG in addition to the ECG, the VCGs were reconstructed from the ECGs and then interpreted by the VCG classification program. The combined ECG and reconstructed VCG results were almost the same as those of the combined ECG and original VCG. Zhou et al. [14] showed that the combination of ECG and synthesized VCG measurements to discriminate between mild RVH with terminal conduction delay and partial RBBB in children significantly improved the classification results. Therefore, it seems interesting to study the combination of ECG and VCG criteria to further improve the performance of PEDMEANS as well.

Finally, the performance of PEDMEANS needs to be evaluated in different clinical settings.

## 7.1 References

[1] Bailey JJ, Berson AS, Garson A, Horan LG, Macfarlane PW, Mortara DW, et al. Recommendations for standardization and specifications in automated electrocardiography: bandwidth and digital signal processing. Circulation 1990;81:730–9.

[2] Dickinson DF. Normal ECG in childhood and adolescence. Heart 2005;91:1626–1630.

[3] Denjoy I, Postma A, Lupoglazoff JM, Vaksman G, Kamblock J, Leenhardt A, et al. Catecholinergic ventricular tachycardia in children. Archives Des Maladies Du Coeur Et Des Vaisseaux 2005;98:506–512.

[4] Garcia-Morales LM, Berber A, Macias-Lara CC, Lucio-Ortiz C, Del-Rio-Navarro BE, Dorantes-Alvarez LM. Use of sibutramine in obese Mexican adolescents: A 6-month, randomized, double-blind, placebo-controlled, parallel-group trial. Clinical Therapeutics 2006;28:770–782.

[5] Hong K, Piper DR, Diaz-Valdecantos A, Brugada J, Oliva A, Burashnikov E, et al. De novo KCNQ1 mutation responsible for atrial fibrillation and short QT syndrome in utero. Cardiovascular Research 2005;68:433–440.

[6] Makarov LM, Kisileva I, Dolgikh V, Bimbaev A, Bairova TA, Drozdova AI. Assesment of parameters of QT interval in children and adolescents. Kardiologiya 2006;46:37–41.

[7] Postma AV, Denjoy I, Kamblock J, Alders M, Lupoglazoff JM, Vaksmann G, et al. Catecholaminergic polymorphic ventricular tachycardia: RYR2 mutations, bradycardia, and follow up of the patients. Journal of Medical Genetics 2005;42:863–870.

[8] Van Leeuwen P, Schiermeier S, Lange S, Klein A, Geue D, Hatzmann W, et al. Gender-related changes in magnetocardiographically determined fetal cardiac time intervals in intrauterine growth retardation. Pediatric Research 2006;59:820–824.

[9] Villain E, Denjoy I, Lupoglazoff JM, Guicheney P. Low incidence of cardiac events with beta-blocking therapy in children with long QT syndrome. European Heart Journal 2004; 25:1405–1411.

[10] Schijvenaars BJ, Kors JA, van Herpen G, Kornreich F, van Bemmel JH. Effect of electrode positioning on ECG interpretation by computer. J Electrocardiol 1997;30:247–56.

[11] Wenger W, Kligfield P. Variability of precordial electrode placement during routine electrocardiography. J Electrocardiol 1996;29:179–84.

[12] Willems JL, Arnaud P, van Bemmel JH, Degani R, Macfarlane PW, Zywietz C. Common standards for quantitative electrocardiography: goals and main results. CSE Working Party. Methods Inf Med 1990;29:263–71.

[13] Kors JA, van Herpen G, Willems JL, van Bemmel JH. Improvement of automated electrocardiographic diagnosis by combination of computer interpretations of the electrocardiogram and vectorcardiogram. Am J Cardiol 1992;70:96–9.

[14] Zhou SH, Liebman J, Dubin AM, Gillette PC, Gregg RE, Helfenbein ED, et al. Using 12-lead ECG and synthesized VCG in detection of right ventricular hypertrophy with terminal right conduction delay versus partial right bundle branch block in the pediatric population. J Electrocardiol 2001;34 Suppl:249–57.

# Summary

The aim of this study was to develop and evaluate a computer program for the interpretation of pediatric ECGs. Earlier the Modular ECG Analysis System (MEANS) for adult ECG analysis had been developed. We now wanted to produce a pediatric version, to be called the PEDiatric Modular ECG Analysis System (PEDMEANS). For that purpose, the measurement part and the diagnostic interpretation part of MEANS had to be modified. The following issues are addressed in this thesis: minimum bandwidth requirements to record pediatric ECGs, estimation of up-to-date normal limits, formalization of expert knowledge with the use of a learning algorithm, and the development and evaluation of PEDMEANS. Finally, the validity of the ECG in diagnosing LVH in children is studied.

Chapter 2 addresses minimum bandwidth requirements for accurate recording of pediatric ECGs. The averaged beats of a large set of ECGs were passed through digital filters with different cutoff points. We measured the absolute errors in maximum QRS amplitude for each simulated bandwidth, and determined the percentage of records with an error greater than 25 μV. We found that in any lead a bandwidth of 250 Hz yields amplitude errors less than 25 μV in more than 95% of the children younger than one year. For older children, a gradual decrease in ECG frequency content was demonstrated. However, we recommend a minimum bandwidth of 250 Hz for the recording of pediatric ECGs. This bandwidth is considerably higher than the previous recommendation of 150 Hz of the American Heart Association.

Chapter 3 presents new normal limits for the pediatric ECG based on a large set of ECGs from children recruited in Rotterdam. These normal limits are used as cutoff points in the diagnostic criteria of PEDMEANS. Normal limits of all clinically relevant ECG measurements were determined for nine age groups and are presented in the form of continuous age-dependent curves. Clinically significant differences were shown to exist in comparison with previously established normal limits. Sex differences could be demonstrated for QRS duration and several amplitude measurements. These new normal limits suggest that diagnostic criteria for the pediatric ECG should be adjusted.

Chapter 4 discusses research that extends our learning algorithm, called EXPLORE (Exhaustive Procedure for LOgic-Rule Extraction), which exhaustively generates rules that fulfill user-specified performance requirements. We present several new techniques to search the feature space more efficiently than by a straightforward, brute-force exhaustive search. The new version of EXPLORE, incorporating these techniques, is able to perform an exhaustive search considerably faster. On five standard data sets from the medical domain the accuracy of the best decision rule generated by EXPLORE is comparable to or surpasses the accuracy of the classifiers generated by the greedy induction algorithms C4.5 and CART.

Chapter 5 describes the modifications made in both the measurement part and the diagnostic interpretation part of MEANS to allow pediatric ECG interpretation. We discuss the use of the learning algorithm in the development process and assess the performance of the system on an independent test set. The performance, on our study population, appears to justify its use in a clinical setting. Preferably, the system should also be evaluated in other clinical centers.

Chapter 6 focusses on the validity of the ECG in diagnosing LVH in children taking the echocardiogram as the reference. LVH in the echocardiogram was defined with the use of a gender-independent LVM index ($LVM_I$). The validity of current pediatric and adult ECG criteria was judged on the basis of an abnormal $LVM_I$, alone and in combination with a LVH score of two pediatric cardiologists based on clinical evidence. In addition, a search was made for possibly better performing combinations of criteria with the use of EXPLORE. We conclude that the ECG criteria have relatively low (<25%) sensitivity rates at 95% specificity in the pediatric hospital population when an elevated $LVM_I$ is taken as the reference for LVH. The performance appears to improve with age. When clinical evidence is also taken into account in the definition of LVH, the sensitivity improved to 43%. By combining two ECG parameters the maximum sensitivity was raised to 32% in the group with abnormal $LVM_I$ and to 51% in the group with a combination of abnormal $LVM_I$ and clinical evidence suggestive of LVH.

Chapter 7 discusses our main findings and for each issue indicates further directions of research. A large data set of ECGs from children with a diversity of cardiac malformations, including rhythm disturbances, is needed to further improve the performance of PEDMEANS. Also, a large database should become available for validation of pediatric ECG analysis systems in general. Furthermore, it seems interesting to study the combination of ECG and VCG criteria to further improve the performance of PEDMEANS which was shown to be of value in adults.

# Samenvatting

Het doel van deze studie was het ontwikkelen en evalueren van een computerprogramma voor de beoordeling van kinder-elektrocardiogrammen. Eerder was reeds het 'Modular ECG Analysis System' (MEANS) ontwikkeld voor de analyse van ECG's van volwassenen. In deze studie wilden we een versie voor kinderen ontwikkelen genaamd het 'Pediatric Modular ECG Analysis System' (PEDMEANS). Hiertoe moest het metingendeel en het interpretatiedeel van MEANS gewijzigd worden. De volgende onderwerpen worden in dit proefschrift besproken: minimum eisen voor de bandbreedte van opname-apparatuur voor het registreren van kinder-ECG's, vaststellen van up-to-date normaalwaarden voor het kinder-ECG, het formaliseren van expertkennis door gebruik te maken van een leeralgoritme, en de ontwikkeling en evaluatie van PEDMEANS. Tot slot is de validiteit van het kinder-ECG voor de diagnose van linker kamerhypertrofie (LVH) onderzocht.

Hoofdstuk 2 bespreekt de minimum bandbreedte eisen voor het nauwkeurig registreren van kinder-ECG's. De gemiddelde slagen van een grote verzameling ECG's zijn gefilterd met digitale filters met verschillende kantelpunten. We hebben de absolute fout in de maximale QRS-amplitude voor elke gesimuleerde bandbreedte gemeten en hebben het aantal metingen met een fout groter dan 25 μV bepaald. We vonden dat in iedere afleiding een bandbreedte van 250 Hz resulteerde in een amplitudefout kleiner dan 25 μV in 95% van de kinderen jonger dan een jaar. Voor oudere kinderen werd een graduele afname in frequentie-inhoud van het ECG aangetoond. Onze aanbeveling is echter om 250 Hz als minimum bandbreedte te hanteren voor het opnemen van kinder-ECG's. Deze bandbreedte is aanzienlijk hoger dan de huidige aanbeveling van 150 Hz van de 'American Heart Association'.

Hoofdstuk 3 presenteert nieuwe normaalwaarden voor het kinder-ECG gebaseerd op een grote verzameling ECG's uit Rotterdam. De normaalwaarden worden gebruikt als drempelwaarden in de diagnostische criteria in PEDMEANS. Normaalwaarden zijn bepaald voor alle klinisch relevante ECG-metingen in negen leeftijdsgroepen en worden gepresenteerd in de vorm van leeftijdsafhankelijke curven. Klinisch significante verschillen zijn aangetoond in vergelijking met eerder gepubliceerde normaal-

waarden. Voor de tijdsduur van het QRS-complex en diverse amplitudemetingen zijn verschillen gevonden tussen jongens en meisjes. Deze nieuwe normaalwaarden suggeren dat diagnostische criteria voor het kinder-ECG aangepast moeten worden.

Hoofdstuk 4 behandelt onderzoek betreffende ons leeralgoritme EXPLORE (Exhaustive Procedure for LOgic-Rule Extraction) dat uitputtend regels genereert die voldoen aan door de gebruiker gedefinieerde prestatievereisten. We presenteren een aantal nieuwe technieken om de zoekruimte efficiënter te doorzoeken en laten zien dat de nieuwe versie van EXPLORE, waarin deze nieuwe technieken zijn opgenomen, een uitputtende zoekactie aanzienlijk sneller uitvoert. Op vijf standaard medische datasets blijkt de nauwkeurigheid van de regels gegenereerd door EXPLORE vergelijkbaar aan of beter dan de nauwkeurigheid van de classificatoren gegenereerd door de 'greedy' leeralgoritmen C4.5 en CART.

Hoofdstuk 5 bespreekt de wijzigingen die gemaakt zijn in het metingendeel en het interpretatiedeel van MEANS om kinder-ECG beoordeling mogelijk te maken. We bespreken het gebruik van het leeralgoritme in de ontwikkeling van PEDMEANS en evalueren de prestatie van het systeem op een onafhankelijke testset. De resultaten op de studiepopulatie lijkt het gebruik in een klinische omgeving te rechtvaardigen. Het is echter aan te bevelen om ook een evaluatie in andere ziekenhuizen uit te voeren.

Hoofdstuk 6 gaat in op de bruikbaarheid van het ECG voor de diagnose van LVH bij kinderen, waarbij het echocardiogram als referentie wordt gebruikt. LVH in het echocardiogram werd gedefinieerd aan de hand van een LVM index ($LVM_I$), onafhankelijk van het geslacht van de patiënt. De validiteit van huidige criteria voor LVH bij kinderen en volwassenen is beoordeeld op basis van een abnormale $LVM_I$, zowel alleen als in combinatie met een LVH-score van twee kindercardiologen op basis van klinische gegevens. Daarnaast is met behulp van EXPLORE onderzocht of combinaties van parameters mogelijk tot betere prestaties leiden. ECG-criteria blijken een relatieve lage sensitiviteit te hebben (<25%) bij 95% specificiteit in een kinderziekenhuispopulatie, wanneer een abnormale $LVM_I$ als referentie voor LVH wordt genomen. De prestaties verbeteren met de leeftijd. Wanneer de aanwezigheid van LVH tevens wordt gebaseerd op klinisch bewijs, verbetert de sensitiviteit tot 43%. Door het combineren van twee ECG-parameters neemt de maximale sensitiviteit toe tot 32% in de groep met abnormale $LVM_I$ en tot 51% in de groep met abnormale $LVM_I$ in combinatie met klinische indicaties van LVH.

Hoofdstuk 7 bespreekt onze belangrijkste bevindingen en toekomstig onderzoek. We menen dat een groot bestand van kinder-ECG's beschikbaar moet komen met een diversiteit aan cardiale afwijkingen, inclusief ritme-afwijkingen, om verdere verbetering van PEDMEANS mogelijk te maken. Bovendien zou een groot bestand ter

beschikking moeten komen voor de validatie van interpretatieprogramma's voor kinder-ECG's in het algemeen. Daarnaast kunnen de prestaties van PEDMEANS mogelijk verder verbeterd worden door ECG-criteria te combineren met criteria op basis van het vectorcardiogram, een combinatie die bij de interpretatie van het volwassen-ECG goed blijkt te werken.

# Curriculum Vitae

Peter Richard Rijnbeek was born on July 29, 1972 in Voorburg. He received his undergraduate education at the Alfrink College in Zoetermeer from 1984 until 1989. He started studying Electrical Engineering at HTS Rijswijk in 1989, where he specialized in Telecommunication Engineering. He received his B.Sc. degree in 1993.

He then continued his study Electrical Engineering at Delft University, where he specialized in Control Engineering. At the same time he worked at the Lung Function Department at the Dijkzigt Hospital in Rotterdam. His graduation project dealt with the development of a model to simulate the human lungs, consisting of a computer controlled pump and airflow resistance. He received his M.Sc. degree in 1996.

In 1996 he joined the Department of Medical Informatics at Erasmus University Medical Center in Rotterdam, where he was involved in research in computerized biosignal interpretation.

Peter Rijnbeek continues his research as a scientific staff member of the Department of Medical Informatics at Erasmus University Medical Center.

# Dankwoord

Samenwerking is cruciaal bij het uitvoeren van wetenschappelijk onderzoek en met name bij het tot standkomen van een proefschrift als dit. Met veel plezier heb ik de afgelopen jaren kunnen samenwerken met een groot aantal collega's in de kindercardiologie en medische informatica. Een aantal wil ik hier in het bijzonder bedanken:

Jan van Bemmel, jouw ruime ervaring in de medische informatica heb ik altijd zeer bewonderd. Ik wil je enorm bedanken voor de ruimte die je me gegeven hebt om naast mijn wetenschappelijke activiteiten ook mijn muzikale talenten te ontplooien.

Jan Kors, als ik van iemand veel geleerd heb de afgelopen jaren dan ben jij het. Het voorbeeld van een perfecte wetenschapper in mijn ogen. Jouw input in dit onderzoek was onmisbaar en ik wil je bovendien enorm bedanken voor alle fijne gesprekken die we hebben gehad over de niet-wetenschappelijke onderwerpen. Ik hoop nog lang met je te kunnen samenwerken.

Maarten Witsenburg, jij was altijd bereid om data ter beschikking te stellen of mee te denken over nieuw onderzoek. Wat ik nooit zal vergeten is de leuke tijd die we hebben gehad in New Orleans en bij het verzamelen van de normale ECG's op de scholen als het 'E-team' in jouw busje.

Gerard van Herpen, veelzijdig en onmisbaar. Jouw hulp bij de interpretatie van de data en met name je ervaring in het schrijven van wetenschappelijke artikelen in de voor mij soms moeilijke Engelse taal, waren bijzonder.

Voor dit onderzoek zijn een groot aantal ECG's verzameld. Ik wil Joke van Woerkom bedanken voor het registreren van de normale ECG's en Marco Kruit en Arno van Vliet voor hun medewerking vanuit kindercardiologie. De interpretatie van de data door John Hess, Andras Szatmari en Derk Jan ten Harkel wordt ook zeer gewaardeerd.

Mama, helaas mocht je het afronden van mijn proefschrift niet meer meemaken. Zonder alles wat ik van jou geleerd heb zou ik dit nooit voor mekaar hebben gekregen. Ik heb dit proefschrift dan ook aan jou opgedragen.