MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-69

One Million Electrocardiograms of Primary Care Patients: A Descriptive Analysis

Emmanuel Chazard[a], Milena Soriano Marcolino[bc], Chloé Dumesnil[a], Alexandre Caron[a], Daniel Moore F. Palhares[c], Grégoire Ficheur[a], Barbara C. A. Marino[c], Maria Beatriz M. Alkmim[c], Régis Beuscart[a], Antonio Luiz Ribeiro[bc]

[a] *Department of Public Health, CHU Lille; EA 2694, University of Lille; Lille, France*
[b] *Medical School, Universidade Federal de Minas Gerais; Belo Horizonte, MG, Brazil*
[c] *Telehealth center, University Hospital of the Universidade Federal de Minas Gerais; Belo Horizonte, MG, Brazil*

Abstract

In 722 cities of Minas Gerais (Brazil), primary care patients can have their ECGs remotely interpreted by cardiologists of the Telehealth Network of Minas Gerais (TNMG), a public telehealth service. As of December 2014, more than 1.9 million ECGs were interpreted. This study analyzed the database of all ECGs performed by the TNMG on primary care patients from 2009 to 2013 (n=1,101,993). Structured patient data and the results of automated ECG interpretation by the Glasgow Program are described. Mean patient age is 51 years old, 59% of them are women. The average body mass index is 25.9 kg/m², with an average increase of 0.15 kg/m² per civil year. Those patients notably have hypertension (33.2%), family history of coronary artery disease (14.5%), smoking (6.9%), diabetes (5.8%), obesity (5.8%) or Chagas Disease (3.0%). Seventy percent of ECGs are normal. This percentage is higher in women (72.3%) and decreases in average by 7.4 every 10 years of life. There are notably 12% of possible myocardial infarction, 10% of possible left ventricular hypertrophy and 8% of possible supraventricular extra systole.

Keywords:

Telemedicine; Electrocardiography; Cardiology; Big Data; Primary Healthcare.

Introduction

The physicians' diagnostic process in primary care is deeply impacted by the representation they form about the prevalence rates of the diseases (pre-test probabilities). Indeed, general practice "has a specific decision making process determined by the prevalence and incidence of illness in the community" [1]. Yet, this decision making process might be biased by the fact that most scientific studies focus on specific populations, especially hospital inpatients, that have increased prevalence rates of severe diseases. To support the diagnostic process in primary care, it is important to publish and discuss descriptive analysis of patients from primary care databases.

The electrocardiogram (ECG) is a widely available method that enables evaluation of the cardiovascular system, to diagnose some diseases or to exclude them. ECGs are easy to perform and available at low-cost. However, their interpretation is not that simple, and in many situations, the general practitioner (or his/her assistant) can perform the exam but is not able to interpret it. For that reason, it has been proposed to have the ECGs interpreted remotely by cardiologists [2]. The first experiment was conducted in 1905, only two years after the first electrocardiograph was ready for use.

To insure access to specialized healthcare to the Brazilian population of remote cities, the State Government of Minas Gerais (Brazil) funded in 2005 the Telehealth Network of Minas Gerais (TNMG) [3]. In this framework, a telecardiology program [4] enables the primary care physicians of remote areas of 722 cities to perform ECGs and transfer them in real time to an analysis center of the TNMG. The ECGs are analyzed by on-duty cardiologists, who send back free-text reports to the physicians the same day. Due to the high number of ECGs analysed per day (on average, 2,200 ECGs), there is a list of ECGs waiting for analysis. A nurse technician coordinates the distribution of the exams among the cardiologists on duty. Emergencies are analysed in less than 10 minutes and routine exams on average in 4 hours. The program has proven to be economically beneficial [5]. More than 1.9 million ECGs have been interpreted in December 2014 [6] but the analysis of this database has still not been published.

Only a few scientific papers provide descriptive analysis of large ECG databases, either as a main or secondary objective, and the results are informative to understand the patients' care and get reliable data about pre-test probabilities of heart diseases [7–9].

The objective of this paper is to provide the community with descriptive statistics about primary care patients and their ECGs in Brazil, by performing a descriptive analysis of the telecardiology database of the TNMG.

Methods

This retrospective observational study assessed the ECG database of the TNMG, which included all the consecutive exams performed from 2009 to 2013 (n=1,101,993).

The ECG database

The ECG database includes all exams performed in the remote health centers from 2009 to 2013, excluding emergency centers (0.2% of the sample). For each exam, the database contains:

- demographic information about the patient: gender, birth date, marital status, educational achievement, income;
- patient's symptoms: pain location (arms, neck, back, precordial, thoracic, epigastric), pain characteristics (caused by effort or emotion, relieved by rest or nitrates, pain intensity), other symptoms (dyspnea, sweating, vomiting, dizziness, syncope, palpitation);
- patient's clinical examination: height, weight, systolic and diastolic blood pressures;

- medications: diuretics, digitalis, beta-blockers, conversion enzyme inhibitors, amiodarone, calcium blockers, others;
- comorbidities or risk factors: arterial hypertension, obesity, diabetes, smoking, hyperlipidemia, personal history of myocardial infarction, personal history of coronary revascularization, family history of coronary disease, Chagas disease, chronic pulmonary disease, chronic kidney disease;
- ECGs represented by a list of time-dependent electric values available in 12 leads;
- administrative information about the ECG analysis: priority level, completion time, time of inclusion in the database, report writing start time, report sending time;
- a free-text report written by the on-duty cardiologist who interpreted remotely the ECG.

Automated computation of descriptive statements by the "Glasgow Program"

The Glasgow 12-lead ECG analysis program is computer software that can automatically interpret ECGs [10]. In this paper, it will be called "Glasgow Program," as in its technical documentation [11]. The Glasgow Program has been evaluated and has achieved very good results for signal processing, e.g., identifying waves and computing axes, durations, amplitudes and intervals [12]. It also obtained acceptable results for rhythm analysis and diagnostic interpretation [13,14]. A review paper concluded in 2001 that the output of computerized ECG processing could be used for epidemiological studies [15].

The Glasgow Program (release 28.4.1, issued on June 16th 2009) was fed with a structured description of the electric signal and a few clinical pieces of information, and was run on the total ECG dataset (n=1,101,993). It was able to output more than 900 different textual messages. We first defined a terminology according to the literature, in order to classify the ECGs the same way it is usually done in the scientific papers [8,9,16–22]. The ECG classification was performed in two steps. First, each ECG was classified into one of the following mutually-exclusive categories: normal, normal variant, abnormal, or fatal technical issue (the ECG cannot be interpreted). The "normal variant" category also contained ECGs that were normal except for rate. Secondly, only in case the ECG could be interpreted, the classification was made of several binary statements (diagnoses) that could be present (with different levels of sureness: "possible", "probable" or "certain") or not, independently from each other.

The Glasgow Program was also able to output quantitative descriptive variables, including heart rate; ventricular rate; average RR with standard deviation; heart rate variability; QRS, P, ST and T frontal axes; P, QRS, and T durations; and PR, QT and ST intervals. The corrected QT interval (QTc) was obtained using the Framingham correction. Those variables were analyzed only for "normal" ECGs, including normal variant.

Statistical analysis

Descriptive univariate statistics were computed from the whole database. For quantitative variables, the mean and standard deviation (SD) were computed. In case of non-normal distribution, quartiles were computed. A histogram was drawn, and/or a density line in case of continuous variable. For qualitative and binary variables, the proportions (prevalence rates) were computed and a bar plot was drawn.

The 95% confidence intervals of means were computed using the normal distribution when appropriate. The 95% confidence intervals of proportions were computed with the Exact Binomial Test, which can get reliable confidence intervals even when the smallest observed count is low or null [23]. The confidence intervals are not reported here when the sample size exceeds 200,000 because they are too narrow.

The Khi² test was used to test the independency between two categorical variables. Student t test or ANOVA was used to test the independency between a quantitative and a binary variable. Generalized linear model was used to test the independency between other kinds of variables. Moving average and linear regression were used to graphically represent the relation between a proportion and a quantitative variable. All the tests were double-sided and interpreted with a 5% significance threshold.

Data management and statistical computations were performed with R statistical computing software [24]. In this paper, when the p value of a test was lower than $2.2e^{-16}$, which is the computation limit of R, we simply wrote "p=0".

Variables having more than 25% missing values were excluded from the analysis. Missing values were studied.

Results

Background and clinical examination of the patients

The patients' mean age is 51 (SD=19.5; 0.5% of missing values). The distribution of the age is displayed on Figure 1. Females comprehended 59.3% of the sample, and this proportion is stable over the years. Patients' mean height is 1.61 m (SD=0.11; 23.7% of missing values). Patients' mean weight is 67.3 kg (SD=16.7; 21.4% of missing values). Patients' mean body mass index (BMI) is 25.9 kg/m² (SD=5.67; 23.5% of missing values; distribution displayed on Figure 2). The BMI is greater than 30 kg/m² in 20.3% of cases. The BMI values increase in average by 0.15 kg/m² per civil year (p=0).

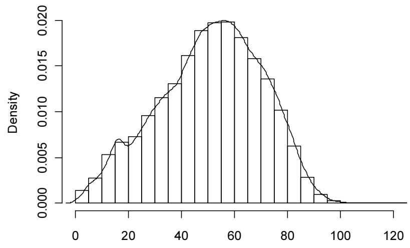

Figure 1 - Histogram of patient's age (years)

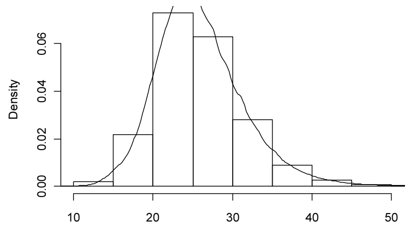

Figure 2 - Histogram of patient's body mass index (kg/m²)

Figure 3 displays the comorbidities and risk factors of the patients. Regarding the comorbidities, 33.2% of the patients have arterial hypertension, 14.5% have a family history of coronary disease, 6.9% are smokers, 5.8% have diabetes mellitus, 5.8% are obese, 3.1% have hyperlipidemia and 3.0%

have Chagas disease. Other statements have a frequency lower than 1%. However, 55.2% of patients are declared not to have any background or identified risk factor. As this part of the form is a set of checkboxes, it is not possible to know the proportion of missing values, but for instance 17.3% of patients having "no obesity" have a BMI greater than 30 kg/m².

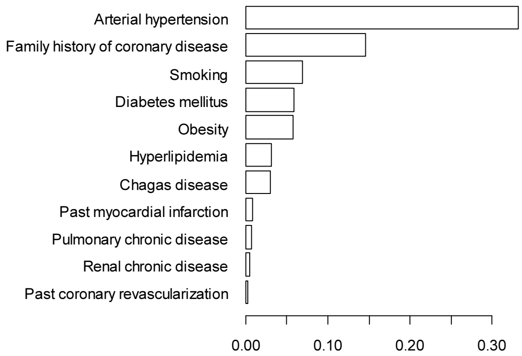

Figure 3 - Comorbidities and risk factors of the patients (prevalence rates)

Regarding the medications, 20.7% of patients take diuretics (the precise class is unknown), 7.9% beta-blockers, 2.8% digitalis, and 0.7% amiodarone. In addition, 95.7% of patients on diuretics suffer from arterial hypertension, and 59.5% of patients with arterial hypertension take diuretics. However, 68.6% of patients declared not to take any drug (see Figure 4). Another drug is taken in 11.0% of cases, but detailed information is only available as free text. As this part of the form is a set of checkboxes, it is not possible to evaluate the proportion of missing values.

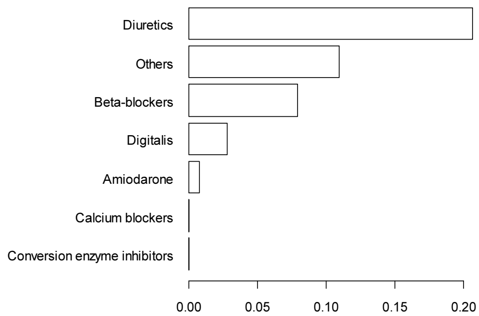

Figure 4 - Medications taken by the patients (prevalence rates)

Administrative management of the exams

At the time of the analysis, the database contains 1,101,993 different exams. The number of exams analyzed per year is displayed on Figure 5. As the year 2013 is incomplete at the time of the data extraction, the extrapolated annual value is represented using a dotted line.

Figure 6 displays the distribution of the time of the day the ECG was performed: 97.4% of them between 7:00 AM and 8:00 PM.

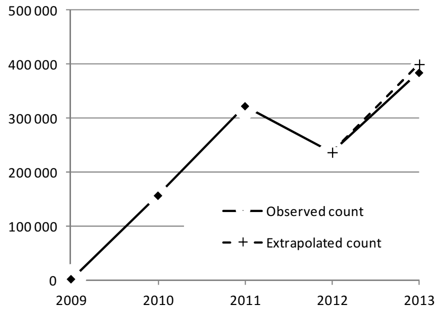

Figure 5 - Number of electrocardiograms performed per year

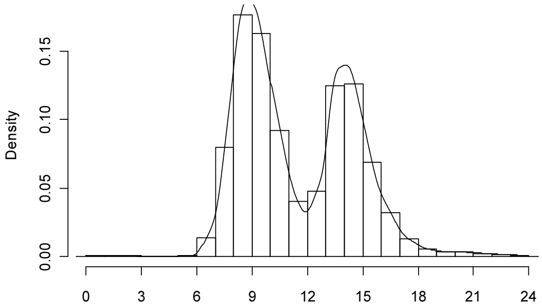

Figure 6 - Histogram of the time of the day the electrocardiogram was performed.

Prevalence rate of normal ECGs

After exclusion of fatal errors, the proportion of normal or normal variant ECGs is estimated to be 69.6%. The percentage of normal ECGs is significantly dependent on the patients' age (p=0), as illustrated on Figure 7. In average, it decreases by 7.43% every 10 years of patient life. The prevalence rate of normal ECGs is lower in men (65.8%) than in women (72.3%) (p=0). It appears to be stable over time (p=11.0%).

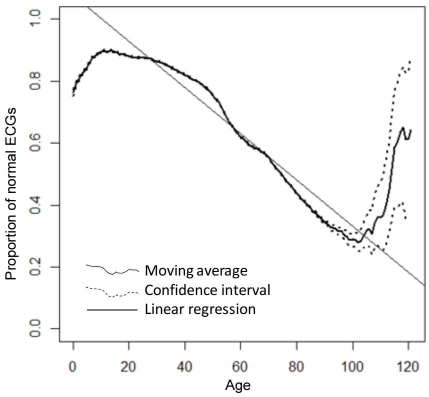

Figure 7 - Proportion of normal electrocardiograms (including normal variant) as a function of the patients' age

Electric description of normal ECGs

The electric parameters of normal ECGs (including "normal variant") are stratified according to gender, and detailed in Table 1. There are 0.00% to 0.86% of missing values for each measurement. The mean of every parameter is significantly different between men and women ($p<10^{-34}$ for each parameter).

Medical description of ECGs

The description of the final statements of the Glasgow Program is displayed in Table 2. The four statements are mutually exclusive. Then, except in case of fatal technical issue, the Glasgow Program can obtain detailed statements that are displayed in Table 3.

Table 1 - Mean (standard deviation) of automatically measured electric parameters of normal electrocardiograms

Variable	Women	Men	t test
Heart rate (BPM)	74.1 (14.8)	70.5 (16.0)	p=0
RR interval (ms)	826 (157)	876 (183)	p=0
PR interval (ms)	152 (28.0)	157 (30.5)	p=0
QRS axis (deg)	28.4 (37.5)	27.2 (44.8)	$p<1e^{-34}$
ST interval (ms)	109 (32.2)	94.0 (31.7)	p=0
QTc interval (ms)[a]	423 (22.8)	414 (24.1)	p=0

(a) with the Framingham correction

Table 2 - Final statement of the Glasgow Program

Final statement	Prevalence rate
Fatal technical issue	0.11%
Normal ECG	33.85%
Normal variant	35.78%
Abnormal ECG	30.25%

Table 3 - Prevalence rates of electrocardiographic statements ("possible", "probable" or "certain" statements)

Statement	Prevalence rate
Global statements	
Non-fatal technical issue	6.28%
Permanent pacemaker	0.45%
Sinus rhythm	93.24%
Non-exclusive ECG abnormalities	
Left ventricular hypertrophy	9.52%
Right ventricular hypertrophy	2.79%
Myocardial infarction	11.61%
Rhythm abnormalities	
Atrial fibrillation or flutter	3.59%
Multifocal or ectopic atrial rhythm	2.06%
Atrial or supraventricular extrasystole	8.21%
Sinusal bradycardia	4.98%
Sinusal or supraventricular tachycardia	3.62%
Accelerated or normal junctional rhythm	0.12%
Idioventricular rhythm	0.19%
Ventricular extrasystole	7.69%
Parasystole	0.00%
Ventricular tachycardia	0.08%
Conduction disturbances	
Sino-atrial block	0.02%
First degree atrioventricular block	5.78%
Second degree atrioventricular block	0.11%
Third degree atrioventricular block	0.15%
Wolff Parkinson White syndrome	0.68%
Left bundle branch block *	5.98%
Right bundle branch block *	4.15%
Descriptive ECG abnormalities	
Repolarization abnormality	39.36%
Bradycardia	5.29%
Tachycardia	3.55%
Short PR	1.99%
Low voltage	1.75%
QRS axis deviation	14.46%
Long QT	2.55%

**: complete, incomplete or fascicular*

Variables with too many missing values

Some variables could not be analyzed due to missing values. Demographic information about the patient included: marital status (48.0% missing), educational achievement (59.2% missing), income (68.8% missing). For those 3 variables, on average, the completeness percentage decreases by 15.4 per year (p=0).

Clinical examination of the patient: systolic and diastolic arterial pressures (respectively 89.1% and 88.7% of missing values). The completeness percentage increases by 1.2 per year (p=0).

Discussion

The objective of this paper was to analyze a million ECGs in a database and provide the community with up-to-date prevalence rates of cardiologic patient conditions in primary care population. This paper shows descriptions of the administrative management of the ECGs (number of exams, hours of performance), the patients (demographic variables, diseases, risk factors and drugs), the ECGs (medical diagnosis obtained from automated interpretation by the Glasgow Program, analysis of the proportion of normal ECGs). This information could participate in the knowledge that constitutes the base of decision-making in primary care.

The proportion of normal ECGs (around 70%) and its relation to the patients' age (it decreases by 7.4% every 10 years of life) are consistent with the literature [8]. The prevalence rates of the diseases are also consistent [8,9].

This study takes advantage of the huge database size that enables the researchers to get precise estimates of prevalence rates and means. However, the analysis raises the issue of completeness and reliability of clinical data. The last section of the results shows that many variables that may be considered secondary by the physicians have a lot of missing values. One can suppose that those values are not missing at random, and may be filled when they support the diagnostic, and left blank in other cases, but this cannot be verified. It is worth noting that the percentage of missing values increases year after year (about +15% per civil year for demographic variables), which probably illustrates the feeling for the physicians that their input is useless and time-consuming. However, this is not in contradiction with Brazilian guidelines for ECG interpretation: the ECG itself does not make the diagnosis. The cardiologist is in charge of interpreting the electric signal, but the clinician has to integrate the clinical symptoms, the risk factors, the personal history of the patient, and other exam results to make the final diagnosis. When the data are provided by the general practitioner, the absence of well shared definitions is probably harmful. In this database for instance, 17.3% of patients with "no obesity" have a BMI greater than 30 kg/m², and are then obese by definition.

The prevalence rates of electrocardiographic statements were obtained from an automated interpretation, and are not as reliable as expert analyses. Additionally, repeated exams from the same patients were not excluded.

Four distinctive characteristics of Brazilian patients can be observed in the analysis results. (1) Three percent of patients suffer from Chagas diseases. (2) Guidelines of the management of patients with arterial hypertension recommend thiazides as a possible first line treatment: 95.7% of patients under diuretics suffer from arterial hypertension, and 59.5% of patients with arterial hypertension receive diuretics (the precise class was not available). (3) The prevalence rate of obesity is high (the average BMI is 25.9 kg/m² and 20.3% of cases are over 30 kg/m²), and increases fast (the BMI values increase in average by 0.15 kg/m² per civil year). (4) Finally, there are only a few smokers (6.9%), but this attribute could have been underreported.

Conclusion

This descriptive analysis of a huge ECG database brings informative results. The analysis may enable primary care physicians to take into account actual prevalence rates of patient conditions that play an important role in the process of medical decision making [1].

Acknowledgments

The authors would like to thank the funding organizations and programs: FAPEMIG (processes #EDT 2372/05 and PPM-00328-08), CNPq (process # 400934/2005-1), ARCUS, the State Department of Health of Minas Gerais (Brazil) and all collaborators from the university centers (UFJF, UFTM, UFU, UFMG, UNIMONTES) and towns involved in the Telehealth Network.

References

[1] The European Definition of General Practice / Family Medicine - Edition 2011 [Internet]. [cited 2014 Dec 18]. Available from: http://www.woncaeurope.org/content/european-definition-general-practice-family-medicine-edition-2011

[2] Hjelm NM, Julius HW. Centenary of tele-electrocardiography and telephonocardiography. J Telemed Telecare. 2005;11(7):336–8.

[3] Alkmim MB, Figueira RM, Marcolino MS, Cardoso CS, Pena de Abreu M, Cunha LR, et al. Improving patient access to specialized health care: the Telehealth Network of Minas Gerais, Brazil. Bull World Health Organ. 2012 May 1;90(5):373–8.

[4] Ribeiro ALP, Alkmim MB, Cardoso CS, Carvalho GGR, Caiaffa WT, Andrade MV, et al. Implementation of a telecardiology system in the state of Minas Gerais: the Minas Telecardio Project. Arq Bras Cardiol. 2010 Jul;95(1):70–8.

[5] Andrade MV, Maia AC, Cardoso CS, Alkmim MB, Ribeiro ALP. Cost-benefit of the telecardiology service in the state of Minas Gerais: Minas Telecardio Project. Arq Bras Cardiol. 2011 Oct;97(4):307–16.

[6] Centro de Telessaúde do Hospital das Clínicas da UFMG [Internet] [cited 2014 Dec 18]. Available from: http://www.telessaude.hc.ufmg.br/

[7] Park MY, Yoon D, Choi NK, Lee J, Lee K, Lim HS, et al. Construction of an open-access QT database for detecting the proarrhythmia potential of marketed drugs: ECG-ViEW. Clin Pharmacol Ther. 2012 Sep;92(3):393–6.

[8] Giuliano I de CB, Barcellos Junior CL, von Wangenheim A, Coutinho MSS de A. Issuing electrocardiographic reports remotely: experience of the telemedicine network of Santa Catarina. Arq Bras Cardiol. 2012 Nov;99(5):1023–30.

[9] Iacoviello L, Rago L, Costanzo S, Di Castelnuovo A, Zito F, Assanelli D, et al. The Moli-sani project: computerized ECG database in a population-based cohort study. J Electrocardiol. 2012 Dec;45(6):684–9.

[10] Macfarlane PW, Devine B, Latif S, McLaughlin S, Shoat DB, Watts MP. Methodology of ECG interpretation in the Glasgow program. Methods Inf Med. 1990 Sep;29(4):354–61.

[11] Physio-Control, Inc., Medtronic B.V. Statement of Validation and Accuracy for the Glasgow 12-Lead ECG Analysis Program [Internet]. 2009 [cited 2014 Dec 18]. Available from: http://www.physio-control.com/uploadedFiles/learning/clinical-topics/Glasgow%2012-lead%20ECG%20Analysis%20Program%20Statement%20of%20Validation%20and%20Accuracy%203302436.A.pdf

[12] Macfarlane PW, Devine B, Clark E. The university of Glasgow (Uni-G) ECG analysis program. Computers in Cardiology, 2005. 2005. p. 451–4.

[13] Salerno SM, Alguire PC, Waxman HS. Competency in interpretation of 12-lead electrocardiograms: a summary and appraisal of published evidence. Ann Intern Med. 2003 May 6;138(9):751–60.

[14] Willems JL, Abreu-Lima C, Arnaud P, van Bemmel JH, Brohet C, Degani R, et al. The diagnostic performance of computer programs for the interpretation of electrocardiograms. N Engl J Med. 1991 Dec 19;325(25):1767–73.

[15] Kors JA, van Herpen G. The coming of age of computerized ECG processing: can it replace the cardiologist in epidemiological studies and clinical trials? Stud Health Technol Inform. 2001;84(Pt 2):1161–7.

[16] Kumar N, Saini D, Froelicher V. A gender-based analysis of high school athletes using computerized electrocardiogram measurements. PloS One. 2013;8(1):e53365.

[17] 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 Aug;26(4):373–8.

[18] 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 Sep;29(4):263–71.

[19] Widman LE, Tong DA. Validation of the EINTHOVEN model-based computerized electrocardiogram rhythm analysis system with three classes of clinical arrhythmias. Am J Cardiol. 1996 Oct 15;78(8):927–31.

[20] Guglin ME, Thatai D. Common errors in computer electrocardiogram interpretation. Int J Cardiol. 2006 Jan 13;106(2):232–7.

[21] Chiu CC, Hamilton RM, Gow RM, Kirsh JA, McCrindle BW. Evaluation of computerized interpretation of the pediatric electrocardiogram. J Electrocardiol. 2007 Apr;40(2):139–43.

[22] Shah AP, Rubin SA. Errors in the computerized electrocardiogram interpretation of cardiac rhythm. J Electrocardiol. 2007 Oct;40(5):385–90.

[23] Clopper CJ, Pearson ES. The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial. Biometrika. 1934 Dec 1;26(4):404–13.

[24] Development Core Team. R: A Language and Environment for Statistical Computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2009. Available from: http://www.R-project.org

Address for correspondence

Emmanuel Chazard, Public Health Department CHRU Lille, 2 avenue Oscar Lambret, 59037 Lille Cedex, France.
Email: emmanuel.chazard@univ-lille2.fr