WORKING PAPER Nr 7 1983

Intraobserver variability of phonocardiogram interpretation.

H.Cleempoel*, P.Mercenier**,
O.Polis*, E.van Thiel*, J.Grosdent**.

* Laboratoire de Cardiologie, Université Libre de Bruxelles (Prof. J.Lequime, Prof. H.Denolin) et Clinique Cardiologique Hôpital Universitaire Saint-Pierre, Bruxelles (Prof. M.Englert)

** Unité de Recherche et d'Enseignement en Santé Publique (Prof. P.Mercenier et H.Van Balen) de l'Institut de Médecine Tropicale, Anvers (Prof. L. Eyckmans)

INTRAOBSERVER VARIABILITY OF PHONOCARDIOGRAM INTERPRETATION

H.Cleempoel [X], P.Mercenier [XX], O.Polis [X], E.van Thiel [X], J.Grosdent [XX].

X Laboratoire de Cardiologie, Université Libre de Bruxelles
(Prof.J.Lequime, Prof.H.Denolin)
et
Clinique Cardiologique Hôpital Universitaire Saint-Pierre,
Bruxelles (Prof.M.Englert)

XX Unité de Recherche et d'Enseignement en Santé Publique
(Prof.P.Mercenier et H.Van Balen) de l'Institut de
Médecine Tropicale, Anvers (Prof.L.Eyckmans)

Correspondance and proofs should be sent to :
Dr. H.Cleempoel
Laboratoire de Cardiologie
Hôpital Universitaire Saint-Pierre,
rue Haute, 322
1000 Bruxelles BELGIQUE

INTRODUCTION

A number of studies have already demonstrated the variability of the interpretation of medical documents submitted on separate occasions to the same observer or to several different observers. It is important for the clinician to reduce such variability to a minimum by identifying the reasons for it and then amending the diagnostic criteria and the conditions of interpretation accordingly.

In the field of cardiology, several studies have been carried out with this object in view (1). Raftery and Holland (2) compared certain components of the clinical examination (cardiac volume and auscultation) performed by five different investigators : two cardiologists, a general practitioner, a gastroenterologist and an epidemiologist. The conclusions of the two cardiologists coincided more frequently than those of the other investigators, but the authors considered that even the conclusions of the cardiologists were too divergent. Similar findings are reported by Taranta et al. (3).

Other authors have reported variability in the palpation of peripheral pulses (4) and the measurement of arterial pressure (5, 6, 7, 8, 9). Inter-observer reproducibility in the interpretation of chest roentgenograms for the diagnosis and quantitative assessment of ventricular septal defect has been studied by Schwarz et al. (10). They obtained a 95% reproducibility result, but higher figures were obtained for the more severe shunts. A number of authors have studied the reproducibility of ECG interpretations (11, 12, 13, 14, 15, 16, 17, 18, 19) and obtained fairly surprising results showing very marked variability in the diagnoses submitted and a lack of reproducibility in general. Two studies have been published more recently on the reproducibility of coronary angiogram evaluation (20, 21).

Most of these studies bear little relation to everyday clinical practice. The authors tended to establish strict

experimental conditions by selecting large homogeneous populations and asking the observers simple questions. Another type of study is the epidemiological survey in which the observer is asked to distinguish between normal and pathological documents obtained from a representative normal population. Some clinical studies concern just one disease and the observer is asked to separate the normal documents from those suggestive of that specific disease, and sometimes to estimate the severity of the lesion.

Such experimental conditions permit a fairly rigorous approach, specially for statistical purposes, but they are far remote from the conditions of everyday clinical activity. The purpose of the work was to see how far the results of these experimental studies would be confirmed in a random sample of documents obtained in the course of routine clinical practice. We tested intra-observer and inter-observer reproducibility and the validity of the interpretation of three documents (chest roentgenogram, ECG and phonocardiogram) which were first read separately and then in combination. The patients included in the study do not represent a selected sample, but the day to day activity, during a period of time, of a cardiology department which carries out pre-operative examinations for congenital and acquired heart diseases.

In our view, the advantage of this methodological approach is that it simulates routine clinical activity in operational conditions. However, the significance of the findings is more open to question because they do not have the same statistical validity as those of epidemiological or clinical studies of larger and more homogeneous populations. In our view, the two types of investigation complement each other, one being closer to laboratory standards and the other to everyday working conditions.

Thus, this work does not claim to solve a problem, but merely seeks to induce the clinician to consider the validity of his approach. Its purpose is not to criticize those who interpret clinical documents or the departments where the documents are produced, but to try and evaluate the extent to

which commonly used clinical methods accomplish what is expected of them. We therefore endeavoured to neutralize as far as possible the "human" variable, in other words, the quality of recording and the competence of the reader. The documents were obtained from a clinical department regarded as eminently reliable and were interpreted by experienced clinicians who are routinely responsible for pre-operative diagnostic examinations. We hope thus to have minimized the element of variability due to the quality of the department and the observers.

However, in an operational study like ours, it is never possible to isolate one variable with as much as in an experimental laboratory study. The results remain, to a certain extent, influenced by external factors impossible to quantify, and comprise therefore a degree of error which is independent from the technique itself.

MATERIAL AND METHODS

A batch of seventy-two successive case files was obtained from a department usually making pre-operative examinations of patients suspected of harbouring congenital or acquired heart diseases. Each patient's documents (chest roentgenogram, ECG, phonocardiogram) were numbered in different order by a nurse who took no part in the evaluation. Only this nurse had access to the chart for identifying the patients.

Each document was submitted on two occasions, separated by a prolonged and variable lapse of time, to two independent and experienced cardiologists who are regularly responsible for making pre-operative diagnostic examinations of patients and who are regarded as top-ranking consultants in their department. Each reading session lasted about two hours during which ten documents were assessed in accordance with an analytical grid worked out with the assistance of the two cardiologists in question. The only information supplied to the observers was the patient's age in each case.

The following factors will be studied :

(a) intra-observer reproducibility for each document in isolation

(b) inter-observer reproducibility for each document in isolation in cases where intra-observer reproducibility was satisfactory.

(c) intra- and possibly inter-observer reproducibility for the three documents considered in combination.

(d) validity of interpretation (measured against true diagnosis) of each document considered separately and marginal utility of these clinical investigations to supplement each other.

Reproducibility is different from validity because two observers may agree on a diagnosis which proves false when confronted with more objective or accurate evidence (catheterization, surgery, necropsy).

We present here the first step of this study.

FIRST STAGE : INTRA-OBSERVER REPRODUCIBILITY OF PHONOCARDIOGRAM INTERPRETATION.

The comparison of two separate interpretations of one document by one observer necessarily constitutes the first step because, apart from providing interesting information, it indicates the diagnoses fo which an inter-observer comparison would be useful. There is obviously no point in comparing reproducibility between two observers if each of them is not even consistent with himself.

The phonocardiogram was generally recorded in three frequency ranges (low, medium, high up to 400 Hz)at the four usual areas, with a simultaneous DII ECG lead (Elema-Schönander with ink recording transducer). The observers were asked to submit a diagnosis, with its degree of probability (certain, probable or possible) and its severity (slight, or significant). (see Annex I).

A diagnosis was classed as "certain" when its probability was considered to be greater than 90% and as "probable" if it lay between 50% and 90%. Fifteen diagnoses were proposed : patent ductus arteriosus, ostium primum defect, ostium secundum defect, valvular pulmonic stenosis, infundibular pulmonic stenosis, mitral stenosis, mitral regurgitation, fixed aortic stenosis, aortic regurgitation, aorta coarctation,

myocardial disorders of uncertain etiology, obstructive myocardiopathies (excluding valvular lesion or shunt), ventricular septal defect, other affections, normal. The observers could reject any tracing on grounds of poor technical quality. There was no guarantee as to the accuracy of the information supplied (recording site, systematic search guided by auscultation, etc.).

The two observers selected 62 phonocardiograms from the original 72 files. They both stressed the limitations of this study due to the often questionable quality of the documents. They pointed out that phonocardiography is an art and that the quality of an evaluation depends closely on the quality of the preliminary clinical examination and the care taken in obtaining the tracing.

So this study makes no attempt to discuss the highly individual "artistic" aspect, which cannot be standardized; it is concerned with routine phonocardiographic technique.

DATA PROCESSING

The recorded data were put into code and transferred to punched cards for the computer. Reproducibility of interpretation was then calculated for each observer according to the probability of his diagnosis (question 1) - "certain", "probable" or "possible" - and the "significant" (question 2a) or "isolated" (question 2b) nature of the presumed disease (see Annex I). For each diagnosis by each of the two observers a table was compiled by the computer according to a standard programme for "all cases", then for "significant" cases and lastly for "isolated" cases (see Table 1, as an example).

RESULTS

All the results for the two observers are shown in Table 2.

Column Q1 shows the figures for intra-observer reproducibility for each of the two observers (O1 and O2) according to the diagnosis and its probability (C = certain, CP = certain or probable, CPP = certain, probable or possible).

Column Q2a shows the corresponding figures for cases where the diagnosis was considered once as "significant" and Column Q2b where the diagnosis was "isolated" (i.e. not associated with other heart diseases).

The denominator represents the number of times the diagnosis was submitted with a certain degree of probability at one of the two readings.

The numerator represents the number of times the same diagnosis was submitted at the other reading with the same degree of probability ("certain", "certain or probable", "certain or probable or possible").

Table 3 resembles table 2 but shows the cases where a diagnosis was submitted as "certain" at one of the two readings and submitted as "certain or probable" at the other reading (see Annex I, questions 3, 4a and 4b).

The results were analyzed by the same procedure for each diagnosis. Reproducibility was calculated for "certain or probable" diagnoses (Question 1 CP). The percentage was then compared with those obtained for the other questions C-C, CPP-CPP, C-CP).

This comparison was performed for the two observers, then the corresponding percentages were calculated for the diagnoses regarded at least once as "significant" or "isolated". We then tried to obtain an overall view which took into account all the implications of the various results. Lastly, we noted the diagnoses which most frequently gave rise to confusion.

1. PATENT DUCTUS ARTERIOSUS

Despite the small number of cases, phonocardiographic diagnosis of patnet ductus arteriosus showed good reproducibility for each of the two observers (4/4 and 4/5), although it was rarely submitted as "certain".

The significant or isolated nature of the diagnosis did not affect reproducibility. There was no confusion with another disease in cases where the diagnosis was submitted as "certain or probable".

This result is expectable; since the patent ductus arteriosus phonocardiogram is very characteristic and specific, it offers reliable confirmation of a clinical impression.

2. OSTIUM PRIMUM DEFECT

This diagnosis was considered in only a very small number of cases; but reproducibility seems none the less to be extremely poor, even in cases where the diagnosis was submitted as significant or isolated.

Confusion occured with ostium secundum defect, patent ductus arteriosus and mitral stenosis.

The criteria used by the clinicians thus seem to be unacceptable; the documents will have to be re-examined and the criteria of analysis discussed again at a later stage.

Admittedly, this is a complex disease which comprises several variables, and this could explain the poor results.

At all events, the phonocardiogram does not seem to be a reliable clinical tool in this disease.

3. OSTIUM SECUNDUM DEFECT

Reproducibility was poor for both observers (7/12 and 6/9), even in cases where the lesion was thought to be significant or isolated.

Confusion occured fairly frequently with ostium primum, ventricular septal defect, mitral stenosis and even aortic stenosis.

This poor result is surprising because the phonocardiographic signs are regarded as characteristic. They are obviously misleading.

The following hypotheses could explain the situation :

a) the phonocardiographic records were not always of the highest quality; for instance, respiratory variations were not always recorded. But the observers did have the opportunity to reject records for technical reasons; what is more, reproducibility was still poor even in cases where the diagnosis was submitted once as "certain".

b) since ostium secundum defect is a fairly frequent disease, it is possible that the observers proposed the diagnosis more readily and accepted the risk of less specificity. This is a fairly common bias which this type of study can indicate.

4. VALVULAR PULMONIC DEFECT

Reproducibility was generally poor (2/4 and 3/4), even where the diagnosis was regarded as significant or isolated. There was confusion with infundibular pulmonic stenosis, ventricular septal defect and even patent ductus arteriosus. This result is also surprising.

The same conclusions and assumptions seem to apply as in the case of ostium secundum defect.

5. INFUNDIBULAR PULMONIC STENOSIS

This diagnosis was considered rarely by observer 1 (1/2), more frequently by observer 2 (4/9). But reproducibility was poor for both observers, even in the cases regarded as significant or isolated.

Confusion occured infrequently with valvular pulmonic stenosis and ventricular septal defect.

6. MITRAL STENOSIS

Reproducibility was excellent for both observers (19/19 and 20/21).

Confusion with ostium secundum defect was very rare. The result is all the more significant because the number of cases in which the diagnosis was considered was high.

The frequent occurence of this disease did not influence the results, probably on account of the highly specific nature of the phonocardiographic signs.

The generally accepted view of the diagnostic value of the phonocardiogram in this disease is thus confirmed.

7. MITRAL REGURGITATION

Reproducibility was generally poor (9/14 and 7/14). But where analysis was confined to diagnoses submitted at least once as significant (5/5 and 7/9) or isolated (6/7 and 5/7), it became excellent, especially for observer 1 who was more strict in his criteria. He submitted the diagnosis less frequently (lower sensitivity), but his interpretation was more often reproducible (greater specificity).

In particular, where observer 1 considered the diagnosis as "certain", he did not change his mind (3/3), unlike observer 2 (3/10).

The criteria of the two observers seem to be different and their varying degree of strictness inevitably affected reproducibility.

The diagnosis of mitral regurgitation was rarely confused with that of ventricular septal defect or ostium secundum defect.

8. FIXED AORTIC STENOSIS

The correlation was fairly good (4/5 and 3/5) but was even better, especially for observer 1, when the analysis was confined to cases regarded at least once as significant (3/3) or isolated (2/2).

Where the diagnosis was submitted at least once as "certain", neither of the observers gave a different answer at the second reading (see Table 3, question 3).

There was no confusion with other diagnoses, except with one case of ostium secundum defect.

10. AORTIC REGURGITATION

Reproducibility was fairly good (6/8 and 5/6), but decidedly better where the analysis was confined to cases regarded at least once as significant (3/3 and 4/5) or isolated (3/3 and 3/3).

Confusion with another diagnosis was rare.

Observer 2 submitted the diagnosis more frequently without impairing the reproducibility of his interpretation.

13. VENTRICULAR SEPTAL DEFECT

Reproducibility was poor and was not improved by limiting the analysis to cases regarded at least once as significant or isolated.

The criteria of the two observers seem inconsistent because, under question 4a - for which reproducibility was perfect - the number of cases concerned was very variable. In this case, a study of inter-observer reproducibility would, no doubt, be interesting and might throw light on the inconsistendy of criteria of the two observers.

9. OBSTRUCTIVE CARDIOMYOPATHIES, 11. AORTIC COARCTATION, 12. MYOCARDIAL DISORDERS OF UNCERTAIN ETIOLOGY

Since these diagnoses occured very rarely, they were not included in this preliminary study.

DISCUSSION

Intra-observer reproducibility is excellent for the diagnoses of patent ductus arteriosus and mitral stenosis. This confirm the general assumption that the phonocardiogram is eminently reliable for the diagnosis of these diseases. But the validity of this assumption should be confirmed by further studies, especially of inter-observer reproducibility, and by the confrontation of interpretations with the definitive diagnoses submitted after catheterization.
Reproducibility is fairly good for the diagnoses of fixed aortic stenosis and aortic regurgitation.

If the analysis is confined to cases diagnosed at least once as significant or isolated, reproducibility improves for mitral regurgitation, fixed aortic stenosis and aortic regurgitation. Here the extent of the lesion and its isolated character seem to play a role. Further studies should focus on such cases.

However, this preliminary study already reveals the need for more severe, more clearly defined criteria of interpretation together with the possible usefulness of two independent readings of each document.

The results are far less encouraging for the diagnoses of ostium primum defect, ostium secundum defect, valvular or infundibular pulmonic stenosis and ventricular septal defect. This throws in question the generally accepted view that the phonocardiogram is a valuable and specific diagnostic instrument in these diseases. Our conflicting findings may be partly explained by the poor quality of the tracing (no systematic record of respiratory variations, for instance). Nonetheless, reproducibility was still poor where a diagnosis was submitted once as "certain". The validity of the diagnostic criteria used by observers for these diseases is thus open to doubt.

Efforts must clearly be made to establish more objective diagnostic criteria for the diseases where reproducibility of interpretation is poor. Perhaps the need for two independent readings should be considered in cases where reproducibility is unsatisfactory.

At all events, our study has served its purpose in questioning the diagnostic criteria used in clinical practice and in showing up the need for greater precision and objectivity. In our view, this kind of critical approach to clinical work should become a permanent feature of the methodology of medical education. Future doctors should be made able to realize that the criteria used in clinical routine are by no means above question (22) and that it is in their interest to analyse, objectivize and improve them.

SUMMARY

Observer variability in the interpretation of the phonocardiogram was studied on the basis of diagnoses submitted by two observers who selected 62 phonocardiograms out of a total set of 72 clinical files.

The patients were either normal or affected with congenital or acquired heart diseases and were being examined with a view to surgery.

The results varied considerably from one diagnosis to another and did not always confirm the generally accepted view of the validity of the phonocardiogram for the diseases considered.

A number of questions must therefore be asked concerning phonocardiogram recording conditions (not sufficiently standardized) and the objectivity of the diagnostic criteria used in clinical practice.

We would like to thank Mr.Hébrant who supervised the analysis of computer results, Mrs.Degimbe who prepared the documents and organized the reading sessions and Mrs. Busine who was in charge of typing.

REFERENCES

(1) Koran L.M. The reliability of clinical methods, data and judgments. N.Engl.J.Med. 1975;293:642-46, 695-701

(2) Raftery E.B., Holland W.W. Examination of the heart : an investigation into variation. Am.J.Epidemiol. 1967;85:438-44

(3) Taranta A., Spagnuola M., Snyder R. Auscultation of the heart by physicians and by computer. In : Enslein eds. New-York : Mc Millan Co, 1964 : 23-52.

(4) Meade T.W., Gardner M.J.,Cannon P., Richardson P.C. Observer variability in recording the peripheral pulses. Br.Heart J. 1968;30:661-65.

(5) Eilertsen E., Humerfelt S. The observer variation in the measurement of arterial blood pressure. Acta Med.Scand. 1968;184:145-57.

(6) Lowe C.R., Mc Keown T. Arterial pressure in an industrial population and its bearing on the problem of essential hypertension. Lancet 1962;1:1086-92.

(7) Labarthe D.R., Hawkins C.M., Remington R.D. Evaluation of performance of selected devices for measuring blood pressure. Am.J.Cardiol. 1973;31:546-55.

(8) King G.E. Taking the blood pressure. JAMA 1969;209:1092-94.

(9) Adler M.W. Measuring blood pressure. Training film to illustrate problems of intra- and inter-observer variation. Brit.J.Med.Educ. 1975;9:195-96.

(10) Schwarz E.D., Dorst J.P., Kuhn J.P., Rowe R.D., Varghese P.J. Reliability of roentgenographic evaluation of ventricular septal defects in children.J.Hopkins Med.J. 1970;127:166-71

(11) Kagan A.R. International comparability in epidemiological studies. 3. Interpretation of electrocardiograms. Milbank Memorial Fund Quart. 1965;43part 2:40-48

(12) Fabian J., Stolz I., Janota M., Konas M. Reproducibility of exercise tests in patients with symptomatic ischaemic heart disease. Br.Heart J. 1975;37:[illegible].

(13) Simonson E.E., Tuna N., Okomato H., Toshima L. Diagnostic accuracy of the vectocardiogram and electrocardiogram. A cooperative study. Am.J.Cardiol. 1966;17:[illegible]74.

(14) Epstein F.H., Doyle J.T., Pollack A.A., Pollack H., Robb G.P., Simonson E. Observer interpretation of electrocardiograms. Suggestions for objective interpretations. JAMA 1961;176:807-50.

(15) Acheson R.M. Observer error and variation in the interpretation of electrocardiograms in an epidemiological study of coronary heart disease. Br.J.Prev.Soc.Med. 1959;13:[illegible]

(16) Davies L.G. Observer variation in reports on electrocardiograms. Br.Heart J. 1958;20:[illegible]

(17) Fischmann E., Cosma J., Diplomyer H.J., [illegible] observer variation of the electrocardiogram. Z.Kreislauf. 1968;75:465-73.

(18) Gorman P.A., Catalayud J.B., Abraham S., Caceres C.A. Observer variation in interpretation of the electrocardiogram. Med.Annals Columbia 1964;33:97-99.

(19) Segall H.N. The electrocardiogram and its interpretation. A study of reports by 20 physicians on a set of 100 electrocardiograms. Canad.Med.Ass.J. 1960;82:2-6.

(20) Detre K.M., Wright P.H., Murphy M.L., Takaro T. Observer agreement in evaluating coronary angiograms. Circulation 1975;52:979-86.

(21) Zir L.M., Miller S.W., Dinsmore R.E., Gilbert J.P., Harthorne J.W. Interobserver variability in coronary angiography. Circulation 1976;53:627-32.

(22) Scheff T.J. Decision rules, types of error, and their consequences in medical diagnosis. Behav.Sci. 1963;8:97-107.

ANNEX I

Question 1 When an observer has submitted a given diagnosis, with a certain degree of probability, at one reading, in what proportion of cases will he submit the same diagnosis and with what degree of probability at the other reading ? For each diagnosis, reproducibility was expressed by three fractions :

- if the reader submitted a "certain" diagnosis (C) on one occasion (denominator), in how many cases did he submit it as "certain" at the other reading (numerator) ?

- if the reader submitted a "certain or porbable" diagnosis (CP) on one occasion, in how many cases did he submit it as "certain or probable" at the other reading ?

- if the reader submitted a "certain or probable or possible" diagnosis (CPP) on one occasion, in what proportion of cases did he submit it as "certain or probable or possible" at the other reading ?

Question 2a Is it possible to improve intra-observer reproducibility by limiting the analysis to diagnoses regarded at least once as "significant" ? Reproducibility was expressed by the same three rates as in the case of global reproducibility.

Question 2b Is it possible to improve intra-observer reproducibility by limiting the analysis to diagnoses regarded at least once as "isolated" ? The analysis was limited to the cases where the diagnosis was not included in an association. For each diagnosis, the figures expressed are as follows : number of cases where the diagnosis was submitted at both readings

number of cases where the diagnosis was regarded as "certain" , "certain or probable", "certain, or probable or possible" at one of the two readings.

Question 3 If the observer sumitted a "certain" diagnosis at one of the two readings, what is the probability of his submitting the same diagnosis as "certain or probable" at the other reading ? Which diagnoses are confused with each other ?

Question 4a Is it possible to improve the reproducibility of diagnoses by limiting the analysis to "certain" diagnoses regarded at least once as "significant" ?

Question 4b Is it possible to improve the reproducibility of the diagnosis by limiting the analysis to "certain" diagnoses regarded at least once as "isolated" ?

TABLE 1

Results of the two independent readings of Observer 1 for the diagnosis of mitral stenosis

		Second reading				
		Cert.	Prob.	Poss.	Excl.	Total
First reading	Cert.	19	0	0	0	19
	Prob.	0	0	0	0	0
	Poss.	0	0	0	1	1
	Excl.	0	0	0	42	42
	Total	19	0	0	43	62

Reproducibility rates

Cert.	19// 19	Pourc. = 100 %
Cert-Prob.	19 / 19	Pourc. = 100 %
Cert-Prob-Poss.	19 / 20	Pourc. = 95 %

Legend

The figures along the horizontal lines represent the number of times a diagnosis was submitted at the first reading and its degree of probability; the figures in the vertical columns relate to the second reading. The results at the bottom of the table therefore express intra-observer reproducibility rates for observer 1 when he diagnosed mitral stenosis as "certain", "certain or probable" or "certain or probable or possible" at the first reading.

TABLE 2

Overall results of intra-observer reproducibility

PHONOCARDIOGRAM		OBSERVER 1			OBSERVER 2		
		Q1	Q2a	Q2b	Q1	Q2a	Q2b
1. Patent ductus arteriosus	C	1/3	1/2	1/3	1/4	1/1	1/4
	CP	4/4	3/3	4/4	4/5	1/2	4/5
	CPP	5/5	3/3	5/5	5/6	2/2	5/5
2. Ostium primum defect	C	0/2	0/2	0/2	0/0	0/0	0/0
	CP	0/2	0/2	0/2	0/1	0/1	0/0
	CPP	0/3	0/4	0/3	0/1	0/1	0/0
3. Ostium secundum defect	C	2/7	2/7	2/7	5/8	5/8	5/7
	CP	7/12	7/12	7/12	6/9	6/9	6/8
	CPP	8/15	8/12	8/14	6/12	6/11	6/10
4. Valvular pulmonic defect	C	0/2	0/2	0/2	1/3	1/3	1/3
	CP	2/4	2/3	2/3	3/4	2/3	3/4
	CPP	3/6	3/3	3/4	4/6	3/3	4/5
5. Infundibular pulmonic stenosis	C	0/1	0/1	0/0	2/2	2/2	2/2
	CP	1/2	1/2	0/1	4/5	3/3	4/4
	CPP	4/10	3/3	3/8	5/13	7/9	5/7
6. Mitral stenosis	C	19/19	15/15	18/18	20/21	16/17	18/19
	CP	19/19	15/15	18/18	20/21	16/17	18/19
	CPP	19/20	15/15	18/19	20/24	16/17	18/19
7. Mitral regurgitation	C	3/3	3/3	3/3	3/10	3/8	3/6
	CP	9/14	5/5	6/7	7/17	7/9	5/7
	CPP	17/26	5/5	7/10	11/24	8/11	6/10
8. Fixed aortic stenosis	C	1/2	1/2	1/2	1/3	1/2	1/3
	CP	4/5	3/3	2/2	3/5	2/4	3/4
	CPP	8/14	3/3	3/6	4/6	3/4	3/4
9. Obstructive cardiomyopathies	C	0/0	0/0	0/0	0/0	0/0	0/0
	CP	0/0	0/0	0/0	0/0	0/0	0/0
	CPP	0/0	0/0	0/0	0/0	0/0	0/0
10. Aortic regurgitation	C	1/4	0/1	1/2	4/6	3/5	3/5
	CP	6/8	3/3	3/3	5/6	4/5	3/5
	CPP	7/9	3/3	3/3	5/9	4/6	3/3
11. Aortic coarctation	C	0/0	0/0	0/0	0/1	0/1	0/1
	CP	0/1	0/0	0/1	1/1	1/1	1/1
	CPP	0/5	0/1	0/5	1/1	1/1	1/1
12. Myocardial disorders	C	0/0	0/0	0/0	0/0	0/0	0/0
	CP	0/0	0/0	0/0	0/0	0/0	0/0
	CPP	0/0	0/0	0/0	0/0	0/0	0/0
13. Ventricular septal defect	C	0/2	0/1	0/2	4/8	3/5	4/7
	CP	6/11	2/4	3/3	5/10	3/4	3/3
	CPP	11/16	4/6	10/13	7/16	3/5	5/14
14. Other	C	0/2	0/2	0/0	0/6	0/6	0/13
	CP	2/6	2/4	1/1 **	2/8	2/3	1/3
	CPP	15/34	4/5	8/20	9/25	4/15	4/10
15. Normal	C	0/1	0/0	0/1	0/0	0/0	0/0
	CP	1/2	0/0	1/2	1/2	0/0	1/2
	CPP	2/3	0/0	2/3	1/3	0/0	1/3

Legend

Q1, Q2a, Q2b see Annex I

C = certain

CP = certain or probable

CPP = certain or probable or possible

TABLE 3

Reproducibility of each diagnosis "certain or probable" at other reading when once diagnosed as "certain".

PHONOCARDIOGRAM	OBSERVER 1			OBSERVER 2		
	Q3	Q4a	Q4b	Q3	Q4a	Q4b
1. Patent ductus arteriosus	3/3	2/2	3/3	4/4	1/1	4/4
2. Ostium primum defect	0/2	0/1	0/2	0/0	0/0	0/0
3. Ostium secundum defect	5/7	4/6	5/7	5/8	4/8	5/7
4. Valvular pulmonic defect	2/2	2/2	2/2	3/3	3/3	3/3
5. Infundibular pulmonic stenosis	1/1	0/0	0/0	4/7	1/1	4/6
6. Mitral stenosis	19/19	15/15	18/18	20/21	16/17	18/19
7. Mitral regurgitation	3/3	3/3	3/3	6/10	6/8	5/6
8. Fixed aortic stenosis	2/2	2/2	2/2	3/5	2/2	3/3
9. Obstructive cardiomyopathies	0/0	0/0	0/0	0/0	0/0	0/0
10. Aortic regurgitation	3/4	1/1	2/2	5/6	4/5	3/3
11. Aortic coarctation	0/0	0/0	0/0	1/1	1/1	1/1
12. Myocardial disorders	0/0	0/0	0/0	0/0	0/0	0/0
13. Ventricular septal defect	2/2	1/1	1/1	3/8	3/3	5/7
14. Other	1/2	0/0	0/0	1/6	1/5	1/3
15. Normal	1/1	0/0	0/1	0/0	0/0	0/0

Legend

Q3, Q4a, Q4b see Annex I