Inter- and Intraobserver Variability in LVH and RVH Reporting in Pediatric ECGs

R. M. Hamilton, M.B.Ch.B., D.Phil.,* K. McLeod, M.D.,† A. B. Houston, M.D.,† and P. W. Macfarlane, D.Sc., F.R.C.P.*

*From the *University Department of Medical Cardiology, Glasgow G31 2ER, UK; and †Department of Cardiology, Royal Hospital for Sick Children, Glasgow G3 8SJ, UK*

Background: Physicians' diagnoses are often used as the gold standard for evaluating computer electrocardiogram (ECG) interpretation programs. As part of a larger study to evaluate the Glasgow pediatric ECG analysis program, inter- and intraobserver variability in the ECG reporting of two pediatric cardiologists was examined.

Methods: The ECGs of 984 children were sent for reporting independently by two cardiologists with all identifying information except age and sex removed. Three hundred twenty ECGs had no clinical indication available, and they were thus reported "blind." For 664 ECGs, the clinical indication was known and included with the ECG trace. All ECGs reported as right ventricular hypertrophy (RVH) or left ventricular hypertrophy (LVH) were returned to the cardiologists without their knowledge for reporting a second time "blind" as to the clinical indication.

Results: When the cardiologists' reports were compared with each other, the provision of clinical information led to greater agreement between them for the diagnosis of LVH (kappa increased from 0.44 to 0.52) but did not substantially affect their agreement in diagnosing RVH (kappa fell from 0.66 to 0.63). Intraindividual comparisons in 166 ECGs revealed that one cardiologist was more consistent in diagnosing RVH and the other more consistent in diagnosing LVH.

Conclusions: This study has demonstrated the difficulties in using cardiologists' diagnoses as the gold standard with which to evaluate pediatric ECGs. **A.N.E. 2005;10(3):330–333**

pediatric; ECG; RVH; LVH; variability

As part of a larger study[1] to evaluate the Glasgow pediatric electrocardiogram (ECG) interpretation program,[2] it was decided to examine the physicians' interpretations of ECGs, since these provided the gold standard with which to evaluate the computer program. Various factors affect physicians' interpretations of adult ECGs including the provision of clinical information[3] and the experience of the clinician.[4] Examination of the peer-reviewed literature failed to elicit any comparable publications on pediatric ECG reporting.

The aim of this study, therefore, was to investigate inter- and intraobserver variability in the interpretation of pediatric ECGs, and whether the variability depended on the provision of clinical information with the ECG.

This study concentrates on the diagnosis of right ventricular hypertrophy (RVH) and left ventricular hypertrophy (LVH). Hypertrophy is one of the most common pediatric ECG diagnoses because heart disease in children tends to result from structural abnormalities with volume and/or pressure overload.

MATERIALS AND METHODS

ECG Recording and Preparation

ECGs were recorded both from inpatients and outpatients at a children's hospital. The ECGs were not selected in any way but 4 patients with ventricular pacemakers and 29 with complex congenital heart disease were excluded (12 had dextrocardia, 12 had univentricular heart, 2 had levocardia with unclear situs, one had mesocardia, one had situs inversus, and one had a crisscross connection). ECGs were recorded on a Siemens Megacart ECG

Address for reprints: P. W. Macfarlane, D.Sc., F.R.C.P., University Department of Medical Cardiology, Queen Elizabeth Building, Glasgow Royal Infirmary, 10 Alexandra Parade, Glasgow G31 2ER, UK. Fax: 0141 552 6114; E-mail: peter.w.macfarlane@clinmed.gla.ac.uk

machine (Drager Medical Systems Inc., Danvers, MA) in the Royal Hospital for Sick Children, Glasgow, and then transferred to a Siemens Megacare ECG management system in Glasgow Royal Infirmary via modem.

Each patient was assigned a study ID number, and all text removed from the ECG trace except the age and sex of the patient. If the clinical indication for the ECG was known, this was copied onto the ECG trace.

ECG Reporting

To test the interobserver variability, ECG traces were sent to two pediatric cardiologists, who reported them independently, each using his/her own preferred criteria. The cardiologists were not provided with any automated measurements or statements. If the ECGs, which were initially reported with a clinical indication, were interpreted as RVH, LVH, or combined ventricular hypertrophy by either cardiologist, they were sent back to both cardiologists to be reported again, this time "blind" as to the clinical indication, thus testing the intraobserver variability. These ECGs were filtered in gradually to the stream of ECGs being sent for reporting without the knowledge of the cardiologists.

Data Analysis

The cardiologists were arbitrarily designated No. 1 and No. 2. For the intraobserver comparisons, each cardiologist's interpretation "blind" was compared with his/her previous interpretation of the same ECG, with the clinical indication given as the gold standard. Any mention of hypertrophy—whether definite, probable, or possible—was taken as "hypertrophy present." "Combined ventricular hypertrophy" was not a separate category for analysis; these cases were considered as having RVH and LVH.

The kappa coefficient was used to measure the extent of the agreement in classification between the two cardiologists.[5] A kappa value of zero means that there is no agreement better than expected by chance, and a value of 1.0 indicates perfect agreement.

RESULTS

Study Group

There were 984 children in the study ranging from 1 day to 18 years; the median age was 4.92 years. Two hundred fifty-one children were aged up to 12 months, there were approximately 50 children in each year of age from 1 to 14 years, and approximately 20 in each year of age from 15 to 18 years. There were 518 males and 466 females. Three hundred and twenty ECGs were reported blind and 664 were reported with the clinical indication. About 75% were cardiology patients and the remainder were hospital inpatients and outpatients from other departments.

Interobserver Comparisons

Table 1 shows the agreement between the two cardiologists in their reporting of ventricular hypertrophy. The provision of clinical information did not greatly change the kappa value for RVH. With LVH, the kappa value increased, i.e., there was greater agreement when the clinical information was given. The kappa values were considerably higher for RVH than for LVH.

Intraobserver Comparisons

One hundred and sixty-six ECGs reported as RVH or LVH by either cardiologist when the clinical indication was given were reported for a second time "blind." In Table 2, each cardiologist's diagnosis of the ECG reported blind is compared with his/her own previous diagnosis of the same ECG

Table 1. Agreement between the Two Cardiologists in the Reporting of RVH and LVH

Method of Reporting by Both Cardiologists[a]	Blind		With Clinical Indication	
Diagnosis by No. 2	RVH	LVH	RVH	LVH
No. of cases (according to No. 2)	21	15	96	38
Kappa	0.66	0.44	0.63	0.52

[a]Both cardiologists reported 320 ECGs blind and 664 ECGs with the clinical indication.
Abbreviations as in text.

Table 2. Comparison of Each Cardiologist's Second Diagnosis (Reported Blind) with His/Her First Diagnosis (With Clinical Indication) of the Same ECG, for a Total of 166 ECGs

ECG Diagnosis Made When Clinical Indication Known[a]	RVH		LVH	
Cardiologist No.	1	2	1	2
No. of cases	90	76	46	35
Kappa	0.53	0.60	0.46	0.72

[a]The ECG diagnosis made with a knowledge of the clinical indication was taken as the gold standard.

reported with the indication given. For both RVH and LVH, cardiologist 2 is more consistent than cardiologist 1, shown by the higher kappa results. Cardiologist 1 is more consistent in diagnosing RVH than LVH; cardiologist 2 is more consistent for LVH than RVH.

DISCUSSION

Interobserver Comparisons

In the interobserver comparisons, the kappa was similar for RVH being reported with and without the clinical indication for the ECG, whereas it increased when LVH was reported with a clinical indication.

It is well known that clinicians' ECG diagnoses can be influenced by clinical information. Hillson et al.[3] found that clinicians receiving clinical vignettes with the ECG trace were more likely to make diagnoses consistent with the vignettes even when these were untrue. Hatala et al.[4] showed a bidirectional effect of clinical history on the accuracy of ECG reports; there was improved diagnostic accuracy when the history suggested the correct diagnosis and reduced accuracy when the history suggested an alternative diagnosis. Provision of a correct history improved the accuracy by 4–12% compared with no history, depending on the interpreter's level of training. Conversely, a misleading history compared with no history reduced the accuracy by 5% for cardiologists, 25% for residents, and 19% for students.

Dunn and Levinson[6] showed that clinical information could influence the ECG interpretation of acute myocardial infarction (AMI). Three cardiologists were each asked to report 52 ECGs blind. Three weeks later, they had to reinterpret the same ECGs but this time the ECGs were accompanied by clinical information, which for half the cases was strongly suggestive of AMI and the other half was mildly suggestive of AMI. Out of the 156 pooled ECGs, 22 changes were made after the clinical information was provided—11 positive changes and 11 negative changes. Changes in interpretation were not affected by whether the ECG was typical for AMI.

The kappa values in Table 1 show that there is more agreement between the two cardiologists on diagnosing RVH than LVH. The low kappa values for LVH imply that the two cardiologists may have quite different criteria for diagnosing LVH. This accords with an observation by Devereux et al.[7] that clinicians may be using additional information in the ECG instead of or in addition to the standard electrocardiographic criteria to diagnose LVH in adults. There may therefore be more variation between clinicians in diagnosing LVH, if these criteria are less widely agreed than the standard ones. Lofsjogard-Nilsson and Nygren[8] suggest that it is hard to select the criteria for diagnosing adult LVH; the same may apply in children.

Disagreements between observers reporting ECGs occur in the adult literature. The reporting of two cardiologists was compared in 381 adult ECGs in a population-based study.[9] They fully agreed in 206 cases, had a minor disagreement in 76, and a major disagreement in 99. The 175 "disagreement" cases were interpreted by a third cardiologist who had a major disagreement with the other two cardiologists in 20 cases. In a study comparing computer interpretations in adults with those of 10 electrocardiographers,[10] there was as much difference between electrocardiographers' interpretations as between those of the computer and the electrocardiographers.

Intraobserver Comparisons

It has been suggested[11] that a 10–20% variability in ECG interpretation may be expected when the same electrocardiographer unknowingly reports an identical ECG on separate occasions. We found greater intraobserver variability than this in our study because there was an added factor of whether or not the clinical information was present.

Table 2 shows that the kappa values for cardiologist 1 are lower than those for cardiologist 2, which implies that cardiologist 1 may have been more influenced by the clinical information than

cardiologist 2. Kappa values for agreement with self shows that No. 1 agrees more for RVH and No. 2 agrees more for LVH.

Use of a Gold Standard

This short article has concentrated on inter- and intraobserver variation. With respect to sensitivity and specificity against a true gold standard, a separate component of the study showed that the cardiologists had a sensitivity of 25% and a specificity of 92% with echo left ventricular mass (LVM) as the gold standard. The computer program had a sensitivity of 14% and a specificity of 96% with respect to the echo LVM.

CONCLUSIONS

This study illustrates the difficulties in using cardiologists' diagnoses as the gold standard with which to evaluate pediatric ECGs. When compared with each other, the cardiologists were more consistent in diagnosing RVH than LVH. For both RVH and LVH, they tended to agree more with each other when provided with clinical information.

In comparing their "blind" reports with their own previous reports of the same ECGs with clinical information provided, one cardiologist was more consistent for RVH and the other for LVH. There was a 56–80% sensitivity level compared to the previous diagnosis of LVH/RVH by the same cardiologist.

Acknowledgments: *We are grateful to the staff of the cardiology department at the Royal Hospital for Sick Children for their assistance with this study.*

REFERENCES

1. Hamilton RM, Houston AB, McLeod K, Macfarlane PW. Evaluation of pediatric ECG diagnosis of ventricular hypertrophy by computer program compared with cardiologists. Pediatr Cardiol 2005 Jan 18 [Epub ahead of print].
2. Macfarlane PW, Coleman EN, Devine B, et al. A new 12-lead pediatric ECG interpretation program. J Electrocardiol 1990;23(Suppl.):76–81.
3. Hillson SD, Connelly DP, Liu Y. The effects of computer-assisted electrocardiographic interpretation on physicians' diagnostic decisions. Med Decis Making 1995;15:107–112.
4. Hatala R, Norman GR, Brooks LR. Impact of a clinical scenario on accuracy of electrocardiogram interpretation. J Gen Internal Med 1999;14:126–129.
5. Altman DG. Practical Statistics for Medical Research. Florida, Chapman & Hall/CRC, 1991, p. 259.
6. Dunn PM, Levinson W. The lack of effect of clinical information on electrocardiographic diagnosis of acute myocardial infarction. Arch Int Med 1990;150:1917–1919.
7. Devereux RB, Casale PN, Eisenberg RR, et al. Electrocardiographic detection of LVH using echocardiographic determination of left ventricular mass as the reference standard. J Am Coll Cardiol 1984;3(1):82–87.
8. Lofsjogard-Nilsson E, Nygren A. Evaluation of LVH as reported from a computer ECG system using echocardiography as a reference. J Electrocardiol 1989;22(Suppl.):179–182.
9. de Bruyne MC, Kors JA, Hoes AW, et al. Diagnostic interpretation of electrocardiograms in population-based research: Computer program research physicians, or cardiologists? J Clin Epidemiol 1997;50(8):947–952.
10. Wolf HK, Gregor RD, Chandler BM. Use of computers in clinical electrocardiography: An evaluation. Can Med Assoc J 1977;117:877–880.
11. Grauer K, Kravitz L, Curry RW Jr, et al. Computerized electrocardiogram interpretations: Are they useful for the family physician? J Fam Pract 1987;24(1):39–43.