Accurate Interpretation of Electrocardiograms by Nonexperts: Validation and Identification of Challenges

Celia Lai1 and Douglas Lee2

1Leslie Dan Faculty of Pharmacy, University of Toronto, Toronto, Canada.
2Institute for Clinical Evaluative Sciences, Toronto, Canada.

Abstract

Objective: This study was primarily designed to validate electrocardiogram (ECG) interpretation via nurse abstractors and computational methodology against the interpretations of an expert electrocardiographer.

Method: The ECGs (n=429) were evaluated independently by 3 different subject groups: a cardiologist, nurse abstractors, and computational software. Required interpretative fields included rhythm, rate, PR interval, QRS duration, type of bundle branch block and type of ischemic changes. The ECG interpretations of the cardiologist served as the comparison standard. A library of synonyms equated computer output with corresponding ECG features of interest to facilitate analysis. Kappa analysis and crude agreement were computed for each interpretative field, comparing cardiologist’s responses against those of nurse abstractors and computational software. Separate sub-group analyses for selective ECG characteristics (paced rhythm, presence of bundle branch block, and number of lead groups with significant Q waves) were conducted.

Results: The validity criteria of this study was determined to be κ > 0.7 or crude agreement >70%. It was found that all nurse abstractors’ interpretations met the validity criteria. The crude agreement ranged from 84.85% up to 100%, while the kappa values spanned over a wider range (0.00 to 0.97). However, nurse abstractor’s ability to detect atrial fibrillation/ flutter was significantly lower when the rhythm was paced –24% lower in the crude agreement between cardiologist’s and nurse abstractors’ interpretations. Except for PR interval (crude agreement = 69.16%, kappa = 0.67), all other computer interpretations met our validation criteria. The crude agreement ranged from 84.85% up to 100%, while the kappa values spanned over a wider range (0.00 to 0.97). However, nurse abstractor’s ability to detect atrial fibrillation/ flutter was significantly lower when the rhythm was paced –24% lower in the crude agreement between cardiologist’s and nurse abstractors’ interpretations. Except for PR interval (crude agreement = 69.16%, kappa = 0.67), all other computer interpretations met our validation criteria.

Conclusions: Aside from PR interval, the relative validity of all other computer software’s interpretations was established. The relative validity of nurse abstractors’ interpretations was also established. This suggests ECG training courses for nurse abstractors can be adapted to focus on their apparent weak areas to further improve their ECG interpretations.

Introduction

Electrocardiography (ECGs) is one of the most useful tools in cardiovascular clinical studies. Kurisu S et al. suggested that electrocardiograms may be useful in predicting short-term prognosis in patients with AMI associated with left main coronary artery [1]. On the other hand, incorrect interpretations of ECGs can carry lethal consequences as Bogun et al suggested that incorrect computerized interpretation of atrial fibrillation that is not corrected by the ordering physicians may result in potentially harmful medical treatments as well as inappropriate use of medical resources [2].

The importance of correct interpretations of ECGs is acknowledged in general, and there have been numerous studies conducted to evaluate competencies of nurses and medical doctors as well as computer software in their ability to correctly interpret ECGs. Salerno et al. conducted a literary review of articles between 1996 to 2002 and concluded that computer software can identify 87% to 100% of various nonrhythmic abnormalities that had been previously identified by an expert electrocardiographers [3]. Salerno et al. also found that noncardiologists identified 87% to 100% of ECGs showing acute myocardial ischemia, correctly classified 72% to 94% of ECGs as meeting criteria for thrombolytic therapy, and diagnosed 95% of ST-segment abnormalities [3].

Although the results obtained by Salerno et al. are useful in preliminary evaluations of the accuracies of ECG interpretations by nonexperts, several subjective aspects and possible contributing factors were not fully considered in many of the consulted studies, and may render inaccuracies to the data. Firstly, the experiences and educational backgrounds of the non-experts enrolled in the studies aforementioned varied. Furthermore, although these studies focused on particular aspects of ECG interpretation, no study has yet investigated whether the presence of paced rhythm and bundle branch block would influence ECG interpretation. In addition, no study has yet looked at whether the presence of significant Q waves in multiple lead groups is associated with the ability by noncardiologists to detect significant Q waves in another lead group (anterior, lateral, anterolateral, or inferior).

With the above considerations in mind, this study aims to further evaluate accuracies of ECG interpretations of nonexperts by: 1) validating electrocardiogram (ECG) interpretations conducted by nurse abstractors and computational analysis against the interpretations of an expert human electrocardiographer, and 2) determining whether certain ECG characteristics (paced rhythm, bundle branch block, and significant Q waves) are subjective to interpretation errors (Table 1).

Methods

Population and Sampling

ECGs collected from the Emergent Heart Failure Study arriving at the Institute for Clinical Evaluative Sciences (ICES) no later than June 12, 2009 (n=429) were evaluated independently by 3 different subject
Table 1: Interpretative fields.

<table>
<thead>
<tr>
<th>Section of Survey</th>
<th>Field Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Section 1</td>
<td>Rhythm, Rate, PR Interval, QRS duration</td>
</tr>
<tr>
<td>Section 2</td>
<td>Bundle Branch Block (BBB)</td>
</tr>
<tr>
<td>Section 3</td>
<td>Ischemic changes in lateral, inferior, anterior &amp; anterolateral lead group: ST elevation, ST depression, Significant Q wave, T inversion</td>
</tr>
</tbody>
</table>

There were 3 sections that were required to be filled out according to the responses received from each subject group (cardiologist, nurse abstractors, and computational software). The interpretative fields in section 1 included: type of rhythm (atrial fibrillation/flutter or sinus), rate, PR interval, QRS duration. Section 2 looked at the presence of bundle branch block (BBB). If a bundle branch block was present, it is further classified into right bundle branch block, left bundle branch block and bifascicular block. Section 3 looked at the presence of ischemic changes (ST elevation, ST depression, significant Q wave, T inversion) in the lateral, inferior, anterior, and anterolateral lead groups. A lead group is indicative of the position of the ischemic changes, and consists of multiple leads. Lead groups are: a) lead V5 & V6 for anterolateral lead group, b) lead 1 and aVL for lateral lead group; in a minimum of 2 out of 3 possible leads (lead I, II, III, aVF); anterior lead group consists of lead V1, V2, V3, V4; and anterolateral lead group consists of leads V5 and V6. There were specific criteria that must be met to satisfy the definition of “significant Q waves”, “ST elevation”, “ST depression”, and “T inversion” in a lead group. The predetermined definition required the corresponding ischemic feature to occur more than once: in both leads (lead 1 and aVL) for lateral lead group; in a minimum of 2 out of 3 possible leads (lead II, III, aVF) for inferior lead group; in a minimum of 2 out of 4 consecutive leads (lead V1, V2, V3, V4) for anterior lead group; and in both leads (lead V5 & V6) for anterolateral lead group.

Measurement of Outcomes

Three sections of interpretative fields were entered into a database according to the responses collected from each subject group (Table 1). Interpretative fields included rhythm, rate, PR interval, QRS duration, type of bundle branch block and type of ischemic changes. Description and definition for each interpretative field were provided to each nurse abstractor in a reference binder during a training session. It should be noted that the definition of ischemic changes (ST elevation, ST depression, significant Q wave, T inversion) in a given lead group implies that the ECG feature of interest appears in at least 2 consecutive leads within the corresponding lead group.

Statistical Procedures

Kappa analysis and crude agreement were computed for each interpretative field comparing cardiologist’s responses against those of nurse abstractors and computer software. The validity criteria of this study was arbitrarily determined to be $\kappa > 0.7$ or crude agreement $>70\%$. These values were assumed to be the minimal degree of validity required for chart abstraction in the Emergent Heart Failure Study. Separate sub-group analyses for nurse abstractors’ versus cardiologist’s interpretations were conducted for the following: 1) paced versus non-paced rhythm; 2) presence versus absence of bundle branch block, and 3) presence of significant Q waves in 0, 1, 2, 2 or 3 lead groups. Statistical significance was set at $P < 0.05$.

Results

All computational interpretations, with the exception of PR interval interpretative, met the validity criteria (Figure 1). The crude agreement and kappa value comparing computer interpretation versus cardiologist’s interpretation for PR interval were found to be 69.16% and 0.67 respectively. On the other hand, all the interpretive fields from the nurse abstractors’ interpretations met the validity criteria (Figure 1). Although the kappa values varied a wide range (0.00 to 0.97), the crude agreement consistently passed the 70% threshold (84.85% to 100.00%) for all interpretation fields. However, nurse abstractor’s ability to detect atrial fibrillation/flutter was lower when the rhythm was paced. In fact, there was a difference of 24% in crude agreement between the responses of the cardiologist and nurse abstractors. Additionally, presence of bundle branch block was associated with a lowered ability for nurse abstractors to detect lateral ST depression and anterior ST depression; with difference in crude agreement of 11.35% and 12.25% respectively. In addition, there appeared to be a trend of lowered crude agreement between the nurse abstractors and the cardiologist group in the detection of significant Q waves in a given lead group when significant Q waves were also present in other lead groups (Figure 2). However, this trend is not statistically significant except for the detection of significant Q waves in the inferior lead group (Figure 2). Specifically, the following observations were made for the crude agreement for inferior significant Q waves: 98.06%
Accurate Interpretation of Electrocardiograms by Nonexperts: Validation and Identification of Challenges

(96.05-99.22) when no significant Q waves were present; 66.67% (52.53-78.91%) when significant Q wave was present in 1 lead group; 25% (5.49-57.19) when significant Q wave was present in 2 lead groups, and 21.43% (4.66-50.80) when significant Q wave was present in 2 or 3 lead groups.

Conclusion & Discussion

Computerized interpretation, except for PR interval, was found to be valid for the purpose of this study. All the nurse abstractors’ interpretations were found to have met the experimental validity criteria. However, there were potential areas that could be improved to increase the accuracy of nurse abstractor’s interpretations. Specifically, nurse abstractors’ ability to accurately interpret ECG decreased under the following situations:

1) detection of atrial fibrillation/flutter with paced rhythm, 2) detection of lateral and anterior ST depression with bundle branch block, and 3) detection of significant Q waves in one lead group when significant Q waves were also present in other lead groups.

This study provided preliminary results to study questions that were not being addressed in current literature. However, there are limitations in the study of ECG interpretations. First of all, the validity criterion was arbitrarily set to meet the minimum demands assumed for most abstraction studies. Nevertheless, these criteria may have to be set higher when ECG readings are used for clinical purposes. Secondly, there were ambiguous terms from computer output that could not clearly be correlated with the ECG interpretation fields used in the study. In addition, low numbers of experimental interpreters for certain sub-group analyses did not provide the power to conclusively shown significant differences.

ECG validation is important because many abstraction studies conducted rely on nurse abstractors for ECG interpretations. In addition, ECG interpretation cannot always be cross-referenced with other medical records in a clinical setting; hence valid interpretation is crucial in maintaining high quality data.

There are several implications that can be drawn from this study. First of all, the shown experimental validity of nurse abstractors’ interpretations suggests that the nurse abstractors are able to independently review ECGs without a cardiologist’s review. This has economical implications as using nurse abstractors to abstract ECG data would help to save financial resources and cardiologists’ time. Secondly, findings from this study can be applied to improve current ECG training courses. It is advised that the content of ECG training courses for nurse abstractors be adapted to focus on their weak areas as discussed to improve the accuracy of ECG interpretations.

Acknowledgement

This project was funded by the Canadian Cardiovascular Outcome Research Team (CCORT) through their Canadian Institutes of Health Research (CIHR) grant program. This study was conducted with study data from the Emergent Heart Failure Study, led by Dr. Douglas Lee at the Institute of Clinical Evaluative Sciences (ICES). Special thanks go to Ms. Alice Chong for assisting with the data analysis, as well as Dr. Douglas Lee and Ms. Audra Stitt for providing invaluable advice.

References


Table 2: ECG interpretations.

<table>
<thead>
<tr>
<th>ECG Interpretation Fields</th>
<th>Computer Interpretation</th>
<th>Nurse Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% Agreement</td>
<td>Kappa</td>
</tr>
<tr>
<td>Sinus Rhythm</td>
<td>91.82</td>
<td>0.8310</td>
</tr>
<tr>
<td>Atrial Fibrillation/Flutter</td>
<td>91.36</td>
<td>0.7887</td>
</tr>
<tr>
<td>Paced rhythm</td>
<td>94.86</td>
<td>0.7783</td>
</tr>
<tr>
<td>ECG rate</td>
<td>94.39</td>
<td>0.9430</td>
</tr>
<tr>
<td>PR interval</td>
<td>69.16</td>
<td>0.6698</td>
</tr>
<tr>
<td>QRS interval</td>
<td>96.50</td>
<td>0.9645</td>
</tr>
<tr>
<td></td>
<td>91.36</td>
<td>0.7384</td>
</tr>
</tbody>
</table>

Section 2

Type of bundle branch block

| Ischemic changes in lateral leads | 78.27 to 99.07 | -0.0318 to 0.0611 | Yes | 84.85 to 100.00 | 0.0000 to 1.0000 | Yes |
| Ischemic changes in inferior leads | 96.26 to 99.96 | -0.0187 to 0.7082 | Yes | 91.61 to 99.07 | 0.0000 to 0.2409 | Yes |
| Ischemic changes in anterior leads | 92.52 to 99.96 | 0.0505 to 0.5863 | Yes | 91.61 to 97.90 | 0.0000 to 0.4709 | Yes |
| Ischemic changes in anterolateral leads | 92.52 to 99.96 | 0.1425 to 0.4537 | Yes | 89.98 to 98.37 | 0.0000 to 0.5017 | Yes |

It can be seen that all of nurse abstractors’ interpretations met the validation criteria. Although the kappa values for all the interpretative fields in section 3 were below the cutoff of 0.7, their crude agreement consistently scored above 70%. The low kappa values could be explained by the low study numbers with the specific ischemic changes in the specified lead groups at any one time. Except for PR interval (crude agreement = 69.16%, kappa = 0.67), all other interpretative fields met the validation criteria. Like the nurse abstractors’ interpretations, the kappa values for all the interpretative fields in section 3 for computational software were below the cutoff of 0.7, yet their crude agreement consistently scored above 70%.
The Department of Cell and Systems Biology provides many exciting possibilities for undergraduate, graduate and post-doctoral education.

Over 30 laboratories conduct research in many areas of modern biology including genomics and proteomics, cell and molecular biology, physiology and developmental biology using plant, animal and microbial systems. See our webpage for further information.